

DOCUMENT RESUME

ED 411 804

IR 056 631

AUTHOR Raitt, David I., Ed.; Jeapes, Ben, Ed.
 TITLE Online Information 96. Proceedings of the International Online Information Meeting (20th, London, England, UK, December 3-5, 1996).
 INSTITUTION Learned Information Europe Ltd., Oxford (England).
 ISBN ISBN-1-900871-04-1
 PUB DATE 1996-00-00
 NOTE 472p.; For individual papers, see IR 056 632-698.
 PUB TYPE Collected Works - Proceedings (021)
 EDRS PRICE MF01/PC19 Plus Postage.
 DESCRIPTORS Access to Information; Computer Software; Cooperation; Current Events; Databases; Distance Education; Electronic Libraries; Electronic Publishing; Foreign Countries; *Futures (of Society); Information Retrieval; *Information Technology; *Internet; Librarians; Online Searching; *Online Systems; Preservation; Search Strategies; User Needs (Information); *Users (Information); World Affairs
 IDENTIFIERS Intelligent Agents; United Kingdom

ABSTRACT

This proceedings volume contains 68 papers. Subjects addressed include: access to information; the future of information managers/librarians; intelligent agents; changing roles of library users; disintermediation; Internet review sites; World Wide Web (WWW) search engines; Java; online searching; future of online education; integrated information education; Internet training programs; distance education; creating a global stock exchange; advocating global forest issues on the Internet; the internet in the Indian context, user needs; new media publishing; pricing models for electronic publishing; information infrastructure; marketing principles of the corporate information center; intranet; electronic current awareness; searching technologies and interfaces; automated patent searching; providing database services; electronic publishing and collection development; public administration information systems; delivery of business information; information on demand; delivery of business news; personally tailored news; imaging; mobile information systems; online scientific and technological information in Nigeria; Internet support for collaborative work; telecommunications; consumer online services in Europe; Minitel; distribution of professional databases on CD-ROM in the Russian information market; online database services in Japan and perspectives on Asia; information as a key resource on the corporate agenda; interactive WWW services; OMNI; online sources for competitor information; services for the digital library; preservation and access to manuscript collection of the Czech National Library; Spanish museum libraries network; web page authoring tools; and resource management in the electronic environment. Author and subject indexes are provided. (SWC)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *



sponsored by

The Guardian

Celebrating 20 years of Online Information

3-5 December 1996

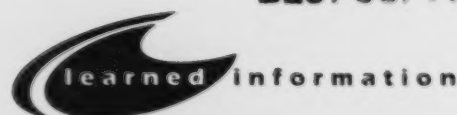
Olympia 2, London, UK

online 96

online information 96
information

Proceedings

BEST COPY AVAILABLE



A Learned Information Event

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☐ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

B.P. Jeapes

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Online Information 96

Proceedings

20th International Online Information Meeting

London 3-5 December 1996

Chairman:

David I. Raitt, The Netherlands

Advisory Committee:

David Bender

Special Libraries Association, USA

Susan Hallam

The Nottingham Trent University, UK

Ruth Handzel

Tel Aviv University, Israel

Barbie Keiser

BEK Inc, USA

Arnoud de Kemp

Springer Verlag GmbH & Co KG, Germany

Cris Kinghorn

Bankers Trust Company, UK

Pier Giacomo Sola

Amitie, Italy

Won, Young-Hee

Korea Database Promotion Center (DPCK), Korea



sponsored by

The **Guardian**

20th International Online Information Meeting

Proceedings

London 3-5 December 1996

Edited by

David I. Raitt

and

Ben Jeapes

Organized by

Learned Information Europe Ltd

*Publisher of Information World Review, The Electronic Library, Online &
CDROM Review, Digital Publishing Strategies, Digital Publishing
Technologies and The Online Manual*



Learned Information Europe Ltd, Oxford

A VNU Company

© 1996 **Learned Information Europe Ltd**
ISBN 1-900871-04-1

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying recording or by any information storage and retrieval system, without permission in writing from the Publisher.

Published by
Learned Information Europe Ltd
Woodside
Hinksey Hill
Oxford OX1 5BE
UK

Typeset by **Learned Information Europe Ltd**

Printed and bound in Great Britain at Unwin Brothers, Surrey



Referee Panel for Contributed Papers at Online Information 96

Armstrong, Chris	Information Automation Ltd, UK
Barker, Philip	University of Teesside, UK
Batt, Chris	Croydon Libraries, Museum & Arts, UK
Citroen, Charles	Delft University of Technology, The Netherlands
Clausen, Helge	State & University Library, Denmark
Day, Joan	University of Northumbria at Newcastle, UK
Dorrington, Linda	Imperial College School of Medicine, UK
East, Harry	The City University, UK
George, Tina	Richbell New Media, UK
Goedegebuure, Ben	FID, The Netherlands
Hallam, Susan	The Nottingham Trent University, UK
Handzel, Ruth	Tel Aviv University, Israel
Heller, Steve	USDA, ARS, USA
Hyams, Peter	Learned Information Europe Ltd, UK
Klasen, Lars	Sema Group InfoData Sweden
Kofoed, Pernille	Statens Information, Denmark
Kosmin, Linda	Jet Propulsion Laboratory, USA
Laitinen, Sauli	VTT Information Service, Finland
Libmann, Francois	FLA Consultants, France
Look, Hugh	Consultant, UK
Mac Morrow, Noreen	University of Strathclyde, UK
McKnight, Cliff	Loughborough University, UK
Mitev, N N	Information Technology Institute, University of Salford, UK
Morris, Anne	Loughborough University, UK
Nieuwenhuysen, Paul	Vrije Universiteit Brussel, Belgium
Oppenheim, Charles	De Montfort University, UK
Owen, Paul	Information Access, USA
Pollitt, A Steven	CEDAR, University of Huddersfield, UK
Retlev, Ulla	Knight-Ridder Information A/S, Denmark
Risseeuw, Mickey	International Translation Centre, The Netherlands
Sanouillet, Anne	URFIST, University of Nice Sophia Antipolis, France
Shimmon, Ross	The Library Association, UK
Spear, Brian	The Patent Office, UK

Stjernvall, Gunnel	Sweden
Stone, Peter	UK
Stubley, Peter	The University of Sheffield, UK
Tuck, Bill	Tuck Segal Associates, UK
van Brakel, Pieter	Rand Afrikaans University, South Africa
Vanouplines, Patrick	Vrije Universiteit Brussel, Belgium
Voigt, Kristina	GSF, Germany
Wallin, Marie	Royal Institute Technology Library, Sweden
Webber, Sheila	University of Strathclyde, UK
White, Martin	TFPL Ltd, UK
Willett, Peter	University of Sheffield, UK
Wilson, T D	University of Sheffield, UK
Wood, Frances	University of Sheffield, UK
Worlock, David R	Electronic Publishing Services Ltd, UK



Contents

Referee Panel	v
Message from the Chairman	xi
Programme	xiii
Rights of access in the Information Society: Europe's quandary over the exploitation of public sector data	
David R. Worlock, <i>European Information Industry Association, UK</i>	1
Information managers/librarians in the year 2006: prophets, princes or Poohbahs?	
Arthur Winzenried, <i>Lilydale Adventist Academy, Melbourne, Australia</i>	7
Information professionals as intelligent agents — or when is a knowbot only a robot?	
Jessie Hey, <i>Multimedia Research Group and Hartley Library, University of Southampton, UK</i>	13
Library users: how they adapt to changing roles	
H. Miido, <i>International Agency for Research on Cancer, France</i>	21
Disintermediation: a disaster or a discipline?	
G. Allen, <i>Foreign and Commonwealth Office, UK</i>	29
The good, the bad and the ugly: Internet review sites	
Alison Cooke, <i>University of Wales, Aberystwyth, UK</i>	
Alison McNab, <i>Loughborough University, UK</i>	
Betsy Anagnostis, <i>Royal Free Hospital School of Medicine, UK</i>	33
Web search engines: key to locating information for all users or only the cognoscenti?	
Nicholas G. Tomaiuolo, <i>Elihu Burritt Library, Central Connecticut State University, USA</i>	
Joan G. Packer, <i>Elihu Burritt Library, Central Connecticut State University, USA</i>	41
Java: an explosion on the Internet	
Tim Read, <i>Sun Microsystems, UK</i>	
Hazel Hall, <i>Queen Margaret College, Edinburgh, UK</i>	49
Twenty years of training in online searching: integrating the Internet with the teaching programme	
Pieter A. van Brakel, <i>Department of Information Studies, Rand Afrikaans University, South Africa</i>	57
The future of online education and training	
Tony G. Rose, <i>Nottingham Trent University, UK</i>	65
Integrated information education as standard part of university curriculum in chemistry	
Jaroslav Silhanek, <i>Institute of Chemical Technology, Czech Republic</i>	71
An iterative development of a training programme in the use of the Internet as an information source and publishing mechanism	
Gobnait O'Riordan, <i>University of Limerick, Ireland</i>	77
Distance teaching in online searching	
Ina Fourie, <i>Department of Information Science, University of South Africa, Pretoria, South Africa</i>	
Dorette Snyman, <i>Library Services, University of South Africa, Pretoria, South Africa</i>	85
Creating a global stock exchange: does the present national system have a future?	
Charles Warner-Allen, <i>Perfect Information, UK</i>	93
Advocating global forest issues on the Internet	
Alois Kempf, <i>Swiss Federal Institute of Forest, Snow and Landscape Research (WSL), Switzerland</i>	99

Internet in the Indian context	
S. Subba Rao, <i>Central Leather Research Institute, Madras, India</i>	107
Looking for the information needle in the Internet haystack	
Helge Clausen, <i>State and University Library, Aarhus, Denmark</i>	115
Missing links — user needs vs. producers' offerings	
Sarah Fahy, <i>J.P. Morgan, UK</i>	125
Final product strategies in purchasing and their impact on the information needs of the organisation	
David Inbar, <i>Minet Information Technologies and Ben Gurion University of The Negev, Israel</i>	129
Chemical structure search on the World Wide Web	
Wolf-D. Ihlenfeldt, <i>Computer Chemistry Centre, University of Erlangen-Nürnberg, Germany</i>	135
Chemical publishing on the Internet: electronic journals — who needs them?	
C.J. Hildyard and B.J. Whitaker, <i>School of Chemistry, University of Leeds, UK</i>	143
The Internet as an information source for environmental chemicals — first results of the evaluation of the metadatabase of Internet resources	
Kristina Voigt, <i>GSF-Forschungszentrum für Umwelt und Gesundheit, Neuherberg, Germany</i>	
Joachim Benz, <i>Universität-Gesamthochschule Kassel, Germany</i>	
Rainer Brüggemann, <i>Institut für Gewässerökologie und Binnenfischerei im Forschungsverbund Berlin, Germany</i>	151
New media publishing	
Ian Ritchie, <i>British Computer Society, UK</i>	161
Creating new pricing models for electronic publishing	
David B. Boelio and Nancy H. Knight, <i>CD-MAX, Inc., USA</i>	167
Information infrastructure: organisational capability for online information provision	
Jane E. Klobas, <i>The Graduate School of Management, The University of Western Australia</i>	173
Marketing principles as applied to the corporate information center	
Suzan A. Brown, <i>Chemical Abstracts Service, USA</i>	181
The intranet: a new concept for corporate information handling	
José Barberá, <i>Fundesco, Madrid, Spain</i>	187
Electronic current awareness in the corporate environment	
Elizabeth Sale, <i>KPMG, UK</i>	195
Intelligent agents: information strategies for the Information Society	
A.J. Garcia-Sierra, <i>BT Electronic Commerce Innovation Centre, Cardiff University of Wales, UK</i> ..	203
Intelligent information harvesting architecture: an application to a high school environment	
Uri Hanani and Ariel Frank, <i>Bar-Ilan University, Israel</i>	211
New searching technologies and interfaces	
Michael L. Weiner and Peter F. Rusch, <i>Manning & Napier Information Services, USA</i>	221
Automated patent searching in the EPO: from online searching to document delivery	
Annemie Nuyts and Charles Jonckheere, <i>European Patent Office, The Netherlands</i>	225
Providing database services in a nationwide research organisation — coexistence of traditional information services and a modern CD-ROM/online hybrid solution	
Benjamin F. Bowman, <i>Information Retrieval Services, Max-Planck-Society, Germany</i>	231
Electronic publishing and collection development, a subscription agent's view	
Philip Wallas, <i>EBSCO Information Services, USA</i>	237
The Internet and local online databases: a novum organum for twenty-first-century library science	
T.D. Webb, <i>University of Hawaii, USA</i>	241

European Public Administration Information Systems on the Internet	
Maria Jesús Fernández Ruíz, <i>Centro de Documentación, Zaragoza, Spain</i>	
José María Angós Ullate, <i>Centro Documentación Científica, Zaragoza, Spain</i>	
Isaac Bolea Aguarón, <i>Random Centro de Informática, Zaragoza, Spain</i>	247
The delivery of business information to the end-user — user perceptions and needs	
Alma Swan, <i>Perspective Business Associates, UK</i>	
Sheridan Brown, <i>Key Consulting, UK</i>	253
Towards real information on demand	
Philip Barker, <i>Human-Computer Interaction Laboratory, University of Teesside, UK</i>	261
New directions in the delivery of business news	
Martin White, <i>TFPL Ltd., UK</i>	271
Tailored news for everyone — the Reuters business briefing case study	
Paul Waddington, <i>Reuters, UK</i>	277
Solving the imaging dilemma for Internet and Intranet applications	
B. Meek, <i>Iterated Systems Incorporated, USA</i>	279
On the move: mobile information systems	
Johan Hjelm, <i>Bonnier Media Lab, Sweden</i>	283
Online scientific and technological information in Nigeria: prospects and possibilities	
R.O. Sodipe, <i>Federal Institute of Industrial Research, Oshodi (FIIRI), Lagos, Nigeria</i>	293
The 'superhighway': a revolutionary means of supporting collaborative work	
C.A. Grant and T.M. Scott, <i>University of Ulster, UK</i>	299
The telecommunications stranglehold on Europe's information use: practical constraints for SMEs and economic assessment based on cost models	
Jean-Claude Delcroix, <i>DECADE sa, Belgium</i>	305
The market prospects for consumer online services in Europe	
Martin White, <i>TFPL Ltd, UK</i>	313
The generation gap: Minitel in the face of the Internet	
Philippe Gilbert, <i>Computer Answer Line, France</i>	321
Some features of the distribution of professional databases on CD-ROM in the Russian information market	
Nickolay Romanov, <i>Association of CD-ROM Users & Producers, Russia</i>	327
Development of online database services in Japan and perspectives on Asia	
Takayasu Miyakawa, <i>Mitsubishi Research Institute, Japan</i>	331
The end-user cometh and cometh again and again	
David Nicholas, <i>Information, Technology, and Change Analytical Research (ITCAR) Group1, City University, UK</i>	
Ingrid Frossling, <i>Department of Economics, University of Stockholm, Sweden</i>	343
Putting information on the corporate agenda as a key resource	
Sylvia E.A. Piggott, <i>InfoPLUS, Canada</i>	349
Disintermediation in the year 2010: using scenarios to identify key issues and relevance of IMPEL2 eLib project	
Graham Walton, Joan Day and Catherine Edwards, <i>University of Northumbria at Newcastle, UK</i>	357
What to do when disintermediation looms	
Cris Kinghorn, <i>Bankers Trust Company, USA</i>	363
Highly interactive WWW services: a new type of information sources	

P. Vanouplines and P. Nieuwenhuysen, <i>University Library, Vrije Universiteit Brussel, Belgium</i>	367
OMNI — alternative approaches to Internet metadata	
Sue Welsh, <i>The OMNI Project, National Institute for Medical Research, UK</i>	379
Smart systems, smart searches	
Bette S. Brunelle, <i>Ovid Technologies, USA</i>	387
Online sources for competitor information	
Arthur Weiss, <i>AWARE, UK</i>	391
Creating services for the digital library	
Dennis J. Crane, <i>Corporate & International Markets, UMI, USA</i>	397
Preservation and access to manuscript collections of the Czech National Library	
Vladimir Karen and Stanislav Psohlavec, <i>Albertina icome Praha s.r.o., Czech Republic</i>	403
Down to earth in cyberspace. Building an electronic library: policy, plans, practice and problems	
Natalia Grygierczyk, <i>Utrecht University, The Netherlands</i>	409
Spanish museum libraries network	
Rosario Lopez de Prado, <i>Museo Arqueologico Nacional Biblioteca, Madrid, Spain</i>	417
Making the best of new information technologies at Eurostat	
Christian Guittet, <i>Commission of the European Communities, Luxembourg</i>	423
View-based searching systems — progress towards effective disintermediation	
A. Steven Pollitt, Martin P. Smith, Mark Treglown and Patrick Braekevelt, <i>The Centre for Database Access Research, University of Huddersfield, UK</i>	433
Web page authoring tools: comparison and trends	
Linda Craney, <i>Prometheus, UK</i>	447
How can documents be put onto the Web?	
Ross Dixon, <i>Pindar PLC, UK</i>	453
Added value in electronic publications	
Theo J. D. Bothma, <i>University of Pretoria, South Africa</i>	459
Resource management in the electronic environment: company library Web pages and collection development principles	
Hazel Hall and Aileen Russell, <i>Queen Margaret College Edinburgh, UK</i>	471
The Internet: the demise of traditional online	
H. Ochsner and K. Thomas, <i>Knight-Ridder Information, Switzerland</i>	477
Author Index	483
Subject Index.....	485



Message from the Chairman David Raitt



Can you really believe that the International Online Meeting has been running — extremely successfully — for the past twenty years? The statistics speak for themselves: nearly three times as many papers presented as at the first meeting in 1977, a similar tripling in the number of delegates, with over 60 countries being represented, and approaching 17,000 visitors to the exhibition. No wonder the conference is widely acknowledged as the single most influential event for the information industry worldwide.

I have had the privilege and honour to be the Conference Chairman for the past fourteen years and have watched the event grow and expand under my guidance. But, of course, this would not have been possible without the foresight of Roger Bilboul, founder and Managing Director of Learned Information, the Organizer, until last year and the support and dedication of LI staff — in particular Jean Mulligan and Elaine Reynolds — as well as all those who have submitted countless papers over the years.

As this was the 20th anniversary, I had hoped (indeed, it was one of the themes in the Call for Papers) that there would be some speakers who would review the achievements and the problems faced over the last twenty years. Curiously there were just a couple of papers which mentioned such developments. Maybe this can be taken as a sign that the information industry is maturing all the time and looking forward to the tremendous changes which are being brought about by the Internet/World Wide Web and consumer online rather than by harking back to the past. Although it must be said that a little more consideration of what was actually done in the past would save some time today. Just as CD-ROMs did not take into account the search and retrieval breakthroughs traditional online had made, so too the WWW has been slow to build on what the pioneers did years ago.

The conference this year offers what I think you will find to be an excellent selection of the many papers submitted (some 180). They cover traditional as well as new topics and they will stimulate thoughts and provoke discussion.

I wish you a good time at the conference and exhibition and a very happy 20th anniversary.

David Raitt



Programme

Tuesday 3 December

Auditorium

09.15-
09.30

Opening Address

The Rt Hon. Roger Freeman MP, Public Services Minister, UK

Auditorium

09.30-
11.00

Chair: David Ralitt, The Netherlands

The Information Society

What is the information society and how will it impact our information industry?

Visions of the Global Information Society and the Implementation process — state of affairs following the first Bangemann report

R F de Brûine, DGXIII/e, European Commission, Luxembourg

Anne C Leer, Oxford University Press, UK

The Labour Party Perspective

Geoff Hoon MP, Shadow Spokesperson on the Information Superhighway and Communications, UK

Rights of Access in the Information Society: Europe's Quandary Over the Exploitation of Public Sector Data

David Worlock, European Information Industry Association, UK

Auditorium

11.10-
12.40

Disintermediation:

The Challenge of Change

Change or become extinct? What are the challenges facing our profession today and will change create new roles for the information professional? Do we take a back seat or should we lead the information revolution into the 21st century? The following session explores the options and offers some solutions/alternative roles for both users and information workers.

Chair: Susan M Hallam, The Nottingham Trent University, UK

Information Managers/Librarians in the year 2006:

Prophets Princes or Poohbahs?

Arthur Winzenried, Lilydale Adventist Academy, Australia

Information Professionals as Intelligent Agents

— or When is a Knowbot Only a Robot

Jessie Hey, University of Southampton, UK

Library Users: How They Adapt to

Changing Roles

Helle Mido, International Agency for Research on Cancer, France

Disintermediation: A Disaster or a Discipline?

Gillian Allen, Foreign & Commonwealth Office, UK

Conference Room 1/2

Web Tools and Developments

The tools necessary for retrieval and promotion on the Web together with the software implications are discussed in this session. From reviews of review services to what tools not to use in the promotion wars, all the latest developments are considered.

Chair: Patrick Vanouplines, Vrije Universiteit Brussel, Belgium

The Good, the Bad and the Ugly: Internet Review Sites

Alison McNab, Loughborough University, UK

Alison Cooke, University of Wales, Aberystwyth, UK

Betsy Anagnostelis, Royal Free Hospital School of Medicine, UK

Web Search Engines: Key to Locating Information for All Users or Only the Cognoscenti?

Nicholas Tomaiuolo, Joan G Packer, Central Connecticut State University, USA

Java: An Explosion on the Internet

Tim Read, Sun Microsystems, UK

Hazel Hall, Queen Margaret College Edinburgh, UK

Developing a Web Site: Some Lessons Learnt

Josy Roberts, Electronic Publishing, Financial Times, UK

Auditorium

14.30-
16.15

Online Education and Training – New Concepts

Training users to get the best from online systems has always been part of the information specialist's brief. But now information can be retrieved from anywhere in the world, manipulated, re-created and sent on elsewhere. Equally, technology has made distance learning a reality. This session considers the current state and future of online education, shows differences in approaches and presents some practical answers to remote training.

Chair: Martha Williams, University of Illinois, USA

Twenty Years of Training in Online Searching: Integrating the Internet with the Teaching Programme

Pieter van Brakel, Rand Afrikaans University, South Africa

The Changing Face of Online Education and Training

Tony G Rose, Hewlett-Packard Laboratories, UK

Integrated Information Education as a Standard Part of University Curriculum in Chemistry

Jaroslav Silhanek, Institute of Chemical Technology, Czech Republic

An Iterative Development of a Training Programme in the use of the Internet as an Information Source and Publishing Mechanism

Gobnait O'Riordan, University of Limerick, Ireland

Distance Teaching in Online Searching

Ina Fourie, Dorette Snyman, University of South Africa, South Africa

Conference Room 1/2

Networked Information Resources

Over the last 20 years technology has been moving the goal posts - the existence of networks, be they Internet, intranet or proprietary, has created unparalleled access to information. The following session highlights some of the latest implications of networking information - from global equity trading to truly collaborative research.

Chair: Terry D Webb, University of Hawaii, USA

Creating a Global Stock Exchange: Does the present national system have a future?

Charles Warner-Allen, Perfect Information Ltd, UK

Advocating Global Forest Issues on the Internet

Alois Kempf, Swiss Federal Institute of Forest, Snow and Landscape Research (WSL), Switzerland

Internet in the Indian Context

Siriginidi Subba Rao, Central Leather Research Institute, India

Looking for the Information Needle in the Internet Haystack

Helge Clausen, State & University Library, Denmark

Auditorium

16.25-
17.30

Producers — The Need to Listen

This session, sponsored by ASIS — European Chapter, provides views from a practitioner and an academic on the needs not being met by producers.

The influence of users' requirements in the formulation of purchasing strategies underlines the importance of producers listening to consumers.

Chair: Theo J D Bothma, University of Pretoria, South Africa

Missing Links: User Needs Vs Producers Offerings

Sarah Fahy, J P Morgan, UK

The Sound of One Hand Clapping – Are Producers Deaf to Users Needs

Derek Law, Department of Information Services and Systems, King's College London, UK

Final Product Strategies in Purchasing and their Impact on the Information Needs of the Organization

David Inbar, Minet Information Technologies and Ben Gurion University of the Negev, Israel

Conference Room 1/2

Chemical Information and the Internet

There is a wealth of chemical information on the Net. Search terms used by chemists, however, are rarely simple and this information can be difficult to find. This session reveals key search techniques and assesses various chemical resources as well as examining the Web's potential for scientific publishing.

Chair: Stephen Heller, USDA.ARS, USA

The Chemical Association Structure Trust Award will be presented at the opening of this session.

Chemical Structure Search on the World Wide Web

Wolf-Dietrich Ihlenfeldt, Universität Erlangen-Nürnberg, Germany

Chemical Publishing on the Internet: Electronic Journals – Who Needs Them?

Benjamin J Whitaker, C J Hildyard, School of Chemistry, University of Leeds, UK

The Internet as an Information Source for Environmental Chemicals – First Results of the Evaluation of the Metadatabase of Internet Resources

Kristina Voigt, GSF-Forschungszentrum für Umwelt und Gesundheit, Germany

Joachim Benz, Universität-Gesamthochschule Kassel, Germany

Rainer Brueggemann, Institut für Gewässerökologie und Binnenfischerei im Forschungsverbund, Germany

Auditorium

09.00-
10.30

Publishing and The New Media - Panel

As digital publishing techniques develop and as alternative media become increasingly available there are both fresh opportunities and threats to established providers. This panel session will in particular debate the possibilities offered by the Internet and how publishers can make money from online publications as opposed to print products.

Chair: Arnoud de Kemp, Springer-Verlag, Germany

Panel members:

Ian Ritchie, British Computer Society, UK

Jeff Meers, IDG Marketing Services, UK

Bela Hatvany, SilverPlatter Information Ltd, UK

David B Boelio, CD-MAX Inc, USA

Bill Thompson, Freelance Writer/Consultant, UK

Tim Frost, Editor, Data Production International, UK

Conference Room 1/2

Corporate Information Management

Information is one of the most important resources for companies today, but it is often undervalued. Giving information the status it deserves involves improved access, organisation and action to promote services at the corporate level.

Chair: Ruth Handzel, Tel Aviv University, Israel

Information Infrastructure: Organisational Capability for Online Information Provision
Jane Klobas, The Graduate School of Management,
The University of Western Australia, Australia

Marketing Principles as Applied to the Corporate Information Center

Suzan Brown, Chemical Abstracts Service, USA

The Intranet: A New Concept for Corporate Information Handling

José Barberá, Fundesco, Spain

Electronic Current Awareness in the Corporate Environment

Elizabeth Sale, KPMG, UK

Auditorium

10.40-
12.10

Information Retrieval and Searching

Software continues to be improved and ways to enhance searching and retrieval, as well as relevance and feedback, are under continual development. Intelligent agents, the quality of information they produce, as well as the impact of this new software on the information society are examined.

Chair: Steven Pollitt, The Centre for Database Access Research, University of Huddersfield, UK

Intelligent Agents: Information Strategies for the Information Society

Adrian Garcia-Sierra, BT Electronic Commerce Innovation Centre, Cardiff University of Wales, UK

Information Harvesting and Intelligent Retrieval Architecture: An Application to High School Environment

Uri Hanani, Bar-Ilan University, Israel

Ariel Frank, Bar-Ilan University, Israel

New Searching Technologies and Interfaces

Michael L Weiner, Peter F Rusch, Manning & Napier Information Services, USA

Automated Patent Searching in the EPO: From Online Searching to Document Delivery

Charles Jonckheere, Annemie Nuyts, European Patent Office, The Netherlands

Conference Room 1/2

Electronic Resources Development and Management

Resources have to be managed and collections have to be developed. What are the principles and strategies for doing this? Taken from academic, suppliers and users standpoints this session represents ways forward for the maintenance of electronic resources.

Chair: Dennis Crane, UMI, USA

Providing Database Services in a Nationwide Research Organization - Co-existence of

Traditional Information Services and a Modern CD-ROM/Online Hybrid Solution

Benjamin F Bowman, Max-Planck-Institute of Biochemistry, Germany

Electronic Publishing and Collection Development, a Subscription Agent's View

Philip Wallas, EBSCO Information Services, USA

The Internet and Local Online Databases: A Novum Organum for Twenty-First-Century Library Science

Terry D Webb, University of Hawaii, USA

European Public Administration Information Systems on the Internet

María Jesús Fernández Ruiz, Centro de Documentación Ayuntamiento Zaragoza, Spain

José María Angós Ullate, Centro Documentación Científica, Universidad de Zaragoza, Spain

Isaac Boleo Aguarón, Random Centro de Informatica SA, Spain

East Hall

14.15-
17.15

Tutorial - sponsored by SLA

Adapting Teambuilding to the New Workplace

This seminar will focus on how the technique of 'teambuilding' can be adapted to whatever your workplace is, whether it is the virtual workplace with people networked together to form a team or a workforce made up of both traditional library staff and the new information professionals.

Speaker: Arlene Farber Sirkin, Washington Resource Consulting Group Inc, USA

Auditorium

14.15-
16.00

New Ways of Delivering Information

Content is everything, but the need to access that content universally is of increasing importance. This session is about the emerging ways that news and business intelligence can today reach individuals both inside and outside the confines of their offices.

Chair: Barbie E Keiser, The College of Insurance, USA

The Delivery of Business Information to the End User - User Perceptions and Needs

Alma Swan, Perspective Business Associates, UK
Sheridan Brown, Key Consulting, UK

Towards Real Information on Demand

Philip Barker, University of Teesside, UK

New Directions in the Delivery of Business News

Martin White, TFPL Ltd, UK

Tailored News for Everyone - The Reuters Business Briefing Case Study

Paul Waddington, Reuters, UK

Conference Room 1/2

Internet Access and Telecoms Aspects

The presentations in this session cover some of the more technical aspects relating to the provision and sharing of data over networks. The effect that the telecomms industry prices have on the free flow of information is discussed, and some solutions proffered.

Chair: Pieter van Brakel, Rand Afrikaans University, South Africa

Solving the Imaging Dilemma for Internet and Intranet Applications

Brian Meek, Iterated Systems Inc., USA

On the Move: Mobile Information Systems

Johan Hjeltn, Bonnier Media Lab, Sweden

On-Line Scientific and Technological Information in Nigeria: Prospects and Possibilities

R O Sodipe, Federal Institute of Industrial Research, Oshodi (FIIR), Nigeria

The Superhighway, A Revolutionary Means of Supporting Collaborative Work

Caroline Grant, Terri Scott, University of Ulster, UK

The Telecommunications Stranglehold on Europe's Information Use: Practical Constraints for SMEs and Economic Assessment based on Cost Models

Jean-Claude Delcroix, DECADE S.A, Belgium and Member of the Board of EIIA

Auditorium

16.10-
17.30

Consumer Online: Realities Behind the Hype

This was meant to be the year that consumer online services finally made a breakthrough outside North America. But what has actually happened, and what lies ahead? How much attention should information professionals pay to such services, and what do the old and the new most have to learn from each other?

Chair: Bill Thompson, Freelance Writer/Consultant, UK

The Market Prospects for Consumer Online Services in Europe

Martin White, TFPL Ltd, UK

The Generation Gap: Minitel in the face of the Internet

Philippe Gilbert, Computer Answer Line, France

Consumer Online - The Pros and Cons for the Information Professional

Martina Flynn, Business and Technical Information Service, University of Limerick, Ireland

Conference Room 1/2

Workshop: Scenario Forecasting and the Role of Business Information

If the business world is to thrive in an ever changing environment it is vitally important to be able to forecast, at least to some degree, what the future may hold. This workshop will show the uses of scenario forecasting as a technique increasingly popular with both companies and large conglomerates.

Speaker: Maria Burke, Manchester Metropolitan University, UK

Auditorium

09.00-
10.30

Disintermediation Revisited: The Human Factor

How can the profession continue to add value in an age of information access for all, and in what ways can new software and communication packages assist in this process? The view of the 'disintermediated' and how they have survived and prospered is also covered.

Chair: Gillian Allen, Foreign and Commonwealth Office, UK

The End-User Cometh and Cometh Again and Again
David Nicholas, Department of Information Science, City University, UK

Ingrid Frossling, Department of Economics, University of Stockholm, Sweden

Putting Information on the Corporate Agenda as a Key Resource

Sylvia Piggott, InfoPLUS, Canada

Disintermediation in the year 2010: Using Scenarios to Identify Key Issues and Relevance of the IMPEL 2 ellb Project

Graham Walton, Catherine Edwards, Joan Day, University of Northumbria at Newcastle, UK

What to do when Disintermediation Looms

Cris Kinghorn, Bankers Trust Company, UK

Conference Room 1/2

Doing Business....in Europe and Asia

This session looks at the supply-side of database services and the shape and size of these developing markets. Facts and figures together with local knowledge give an insight into new and 'old' markets, creating a basis for marketing online products in 'less developed' countries. Some original research will be announced.

Chair: Pier Giacomo Sola, Amitie, Italy

Online Business Information Market in Europe - its value shape and future

David Mort, IRN, UK

Jill Cousins, Learned Information Europe Ltd, UK

Anthony Easton, Reuters Business Information, UK

Some Features of the Distribution of Professional Databases on CD-ROM in the Russian Information Market

Nickolay Romanov, Association of CD-ROM Users & Producers, Russia

Development of Online Database Services in Japan and Perspectives on Asia

Takayasu Miyakawa, Mitsubishi Research Institute Inc., Japan

Auditorium

10.40-
12.10

Nuggets on the Net

This session discusses experiences of trying to source valuable information on the Internet, and strategies for sifting gold from the surrounding dross.

Chair: Heide Clausen, State University Library, Denmark

Highly Interactive WWW Services: A New Type of Information Sources

Patrick Vancouplines, Paul Nieuwenhuysen, Vrije Universiteit Brussel, Belgium

OMNI - Alternative Approaches to Internet Metadata

Sue Welsh, National Institute for Medical Research, UK

Smart Systems, Smart Searches

Bette S Brunelle, Ovid Technologies, USA

Online Sources for Competitor Information

Arthur Weiss, AWARE, UK

Conference Room 1/2

Electronic Libraries and Services

The Internet and new digital publishing techniques are providing wonderful opportunities for libraries to become more creative in making their collections more widely available and accessible.

The Electronic Library Best Paper Award will be presented at the opening of the session.

Chair: David Worlock, European Information Industry Association, UK

Creating Services for the Digital Library

Dennis Crane, UMI, USA

Memory of the World and the World Wide Web -

The Czech National Library Digitised

Vladimir Karen, Stanislav Pschlaivec, Albertina icome Praha sro, Czech Republic

Down to Earth in Cyberspace. Building an Electronic Library: Policy, Plans, Practice and Problems

Natalia Grygierczyk, Utrecht University, The Netherlands

Spanish Museum Libraries Network

Rosario López de Prado, Museo Arqueológico Nacional, Spain

Auditorium

14.00-
15.30

Creating Exciting Pages and Documents

It is clearly important for Web pages to be designed for maximum impact, so knowing the best tools for the job is vital. This session provides practical advice as well as informed suggestions as to how you should publish on the Web.

Chair: Alison Cooke, University of Wales, Aberystwyth, UK

Web Page Authoring Tools: Comparison and Trends
Linda Craney, Prometheus, UK

How Can Documents be put onto the Web?
Ross Dixon, Pindar plc, UK

Added Value in Electronic Publications
Theo J D Bothma, University of Pretoria, South Africa

Resource Management in the Electronic Environment: Company Library Web Pages and Collection Development Principles
Hazel Hall, Aileen Russell, Queen Margaret College
Edinburgh, UK

Conference Room 1/2

The Effects of New Technology

New technology can enhance an organisation's products and services and safeguard its property. But will it also make the searcher's life easier, or will it constrict? Leading-edge case studies are included in this session.

Chair: Philip Barker, University of Teesside, UK

Hybrids and the Internet

Mark Brenner, Editor, Business Computing Monthly, UK
Marika van der Donk, VNU, UK

Making the Best of New Information Technologies at Eurostat

Christian Guittet, Commission of the European Communities, Luxembourg

View-based Searching Systems - Progress Towards Effective Disintermediation

A Steven Pollitt, **Martin P Smith**, **Mark Treglown**,
Patrick Braekeveld, The Centre for Database Access
Research, University of Huddersfield, UK

Auditorium

15.40-
17.00

The Future - Is the Past Any Guide?

The Internet has undeniable benefits for online searchers but what are the repercussions for the industry? The current state of the online industry will be considered and questions on the future of traditional online hosts and services will be posed.

Chair: David Raitt, The Netherlands

The Internet: The Demise of Traditional Online

Heinz Ochsner and **Karen Thomas**, Knight-Ridder Information, Switzerland

Panel discussion reflecting on how the industry has come of age and the way ahead into the next millenium.

Larry Rees, M.A.I.D, UK

Barry Mahon, EUSIDIC, Luxembourg

Ronald Dunn, Information Industry Association, USA

Paul Blake, IWR, UK



Rights of access in the Information Society: Europe's quandary over the exploitation of public sector data

David R. Worlock

European Information Industry Association, UK

Abstract: *The paper reviews ten years of debate in Europe on commercial rights of access to government information, starting with the UK's innovative Government Tradable Information Guidelines of 1996 and ending with the European Commission's Draft Green Paper on Rights of Access in 1996. US market development and commercial rights of access are compared to the European situation, and the needs of the European information industry in this regard are reviewed in detail. The paper looks at access to legal and business information as critical case studies, citing access to the UK VAT Register as a recent case study. The decline in government copyrights and the increase in public domain data released onto the Internet is contrasted with the growing tendency to create new information monopolies through privatisation and deregulation.*

Keywords: electronic democracy, copyright, public sector, data protection, privacy, electronic publishing, European Union

In the 10 years of active debate on commercial rights of access to public sector information in Europe, the situation has changed entirely and yet remained very much the same. It is possible for private sector activists to say, as one did recently, 'Government information is free, yet everywhere it is in chains', while a public servant could riposte 'Government information was in chains but now most of it is on the Internet'. Yet this debate was intended to provide a stimulus for the growth of the information marketplace and of the commercial sector in electronic publishing in Europe by emulating US experience and by providing a form of developmental support without subsidy. Government, by providing primary information without charge, would stimulate inventive entrepreneurs to add high levels of value in order to make such information more effectively available to targeted groups of users. This notion of value added was a vital one in the ideology of the private sector: it allowed companies to live with the idea of non-exclusive licensing by government to the private sector, since value added would be the competitively exclusive element, and it allowed government at once to licence the core information non-exclusively and add value in its own context. This meant that the idea of non-exclusive access to core or primary information was not at odds with a government printer, like the HMSO in the UK, adding its own levels of value to government information for re-sale, or business information parastatal organisations like CERVED or CAMERDATA adding value to information for their user communities, or the German organisations like the FIZ institutes or STN network adding value to their own databases by placing them in an appropriate retrieval and networking environment.

The essential problems with this apparently felicitous market development model emerged in practice. It proved extremely difficult to define source or primary information, to define value-added services and, indeed, to get governments to keep their side of the bargain and set up appropriate licensing procedures. Such raw primary data, which information market companies had simply seen as core statistical data, government announcements and administrative orders, statutes and secondary legislation, and databases of information collected from the public at large by reason of statutory requirement, have often proved to be difficult to track down in practice. Both government and the private sector have been guilty of deluding themselves that there is more value in these databases than there sometimes is, and where genuine attempts have taken place to catalogue government information holdings (the UK Department of the Environment Spatial Database Initiative is a case in point), private sector take-up has been disappointing. Where such take-up has taken place, negotiations have been drawn out into exhaustive contractual debates which frustrate the essential purpose behind the original UK guidelines of 10 years ago.

Many of the instances given in this paper concern practice in the UK. This is deliberate, on two grounds. In the first instance, the UK Guidelines on Government Tradable Information of 1986 were the first development in Europe of this nature, and much of the thinking from that process was adopted into the European Commission's Guidelines on the Synergy between the Public and the Private Sector of 1989. On the other hand, the UK's retention and use of Crown copyright demonstrates attitudes towards retention of control of government information which are amongst the most clear-cut in Europe and which tend to throw up case studies of real interest. This does not imply for a moment that issues discussed are not fully relevant in all other parts of the European Union, or indeed that UK practice, because it is so obvious, presents a particular problem. Unwillingness to

Online Information 96 Proceedings

Page 1

BEST COPY AVAILABLE

licence utilisation of government data does not need legal sanction to be effective. The number of instances in which database developers have been able to collect information from all Member States in order to create integrated, cross-border, value-added products which help and support trading activity and derive from non-exclusive licences granted by the Member State governments, remain very few and far between. In some Member States it is clearly and legendarily impossible for commercial entities in other Member States even to attempt to licence government information, regardless of the equality of treatment issues which may be involved.

Let us go back, however, to the beginning and revisit some of the underlying principles to see what practical sense we have made of them over the last ten years. Pioneers of access to government information arguments in Europe 10 years ago did not think that they were making particularly revolutionary suggestions. They believed that if the strategic goal was to create a free trade area in Europe with greater transparency and lower barriers, in order to promote and increase trade and therefore wealth and employment on a transnational basis, then the removal of barriers to trade caused by inadequate, insufficient or unreliable information was of prime importance. They identified government as the source, although not the only source, of information used in the trading process and, at a time when the prevailing notion was one of 'level playing fields', it seemed natural to seek common conditions of access to government information in the various states of the European Union in order to create conditions sympathetic to the development of pan-European databases. Insofar as the *Guidelines on the Synergy of Public and Private Sectors in the Information Marketplace* were an attempt to do this, they are commonly reckoned to have failed. But the key markers that they laid down were as follows:

1. Governments should licence freely and non-exclusively to allow the information industry to grow through adding value through re-use of public sector database collections.
2. Governments should be encouraged to create catalogues of licensable information and make them available to private sector developers.
3. Having licensed private-sector companies to invest in developing government data to serve markets better, governments should not normally go out and create other services with the tax-payers' money in order to compete with these licences.
4. Re-use of primary source data by private sector organisations should normally be free, although the Guidelines recognised the rights of governments in some instances to make cost recovery of the cost of actually supplying the information to the private sector (as distinct from the costs of its collection).
5. Article 2(4) of the Berne Convention should be utilised to allow governments to 'exempt from copyright all texts covering legislation, public administration, economic and social policy and activity, and norms and standards'.

These principles remain close to the heart of the private sector database industry in Europe and, indeed, are the most important elements within the 19 articles which the European Information Industry Association has put forward to the European Commission as part of its expectations for a draft directive on commercial rights of access to government information.

With guidelines like this in place (and the UK national version even had a model contract produced by HMSO attached to it), it might be thought that matters would change rapidly. They did not, and it is instructive to see why. We identify the following major reasons:

1. The guidelines were not effectively communicated to responsible individuals within Member State governments and, indeed, few Member State governments had officials deployed specifically to manage information policy. Where such policy coordination did exist, as for a time in the UK, it was simply coordination and not control. It did not, for example, prevent wayward departments from going in their own directions.
2. Neither government nor the private sector were sufficiently aware of government primary source data holdings. No catalogues were produced in many countries and, where they were (as indicated above, in the UK), they were ineffectively deployed.
3. In the late 1980s Member State governments were becoming very concerned to reduce government expenditure in the face of industrial recession and began to see government information as a revenue source. Just as the private sector over-emphasised the economic opportunity on its side, so the public sector did the same. Hard-pressed bureaucrats sought budget relief from data sales, even though any revenue flows tended to go to central Treasury controls rather than to individual operating departments.
4. During this period, and increasing in strength throughout the period, the pervasive movement towards privatisation of state functions and the removal of former government activities into parastatal or privatised organisations was a key policy feature. There was a tendency for organisations to be privatised with their information resources. British Telecom took the United Kingdom telephone directory with it, for example, and certainly regarded this as a commercially exploitable entity (selling annual licences for re-use for sums reputed to be in excess of £500,000 per annum) until its regulator proposed re-regulating this area in 1995.

There are countless other examples of this unthinking creation of information monopolies by government, and a certain amount of evidence that where government was conscious that it was creating new monopolies, it succumbed upon the grounds that the information concerned was a vital part of the asset base of the newly privatised organisation and could be one of its exploitable benefits. Besides taking more and more government servants off the payroll, the privatisation movement has sharply reduced the amount of information within the scope of the guidelines, and this in itself might be taken as a critical reason for their failure.

Online Information 96 Proceedings

Page 2

It is necessary to give this background in order to demonstrate why we are where we are. However, we need to keep on reminding ourselves that the arguments originally advanced in the guidelines had nothing to do with government policy or making better informed citizens. The arguments put forward then, as now, were commercial and economic, and they related to developmental opportunities which were as clearly in front of European Union citizens then as they are now. And since some of the opposing interests have at times expressed a view that this simply reflects the avaricious desires of private-sector information service developers who wish to corner public-sector information and re-sell it at a profit to tax-payers, who funded its collection in the first place, it is worth looking at some of the areas of opportunity that are created if access to government information without barriers is allowed:

- The potential for cross-border information services is enhanced. Private sector investment in the translation and harmonisation of information resources becomes more realistic if the rules for release of information are the same in each Member State, the point of availability is clearly known, the type of cataloguing that we have referred to above has taken place to allow relevant resources to be recognised, and the costs of obtaining the information are calculated on the same basis in each case;
- The opportunity is enhanced in network environments where information service developers are able to charge their customers not for the underlying information content but simply for the value-added service superstructure located on top of it. Let us postulate, for example, a service that improved the European Commission's existing TED (Tenders Electronic Daily) service, and which drew deeply upon regional government tendering opportunities throughout Member States. The added service value could be intelligent agent applications which allowed users to identify subject areas and be notified by the system when opportunities of interest to them were lodged on the database. Without arguing the virtues or lack of them of such a system, the point to be made is simply that current networking styles on the Internet create conditions like this where underlying information may be free but users may choose to access information through a paid-for service that adds value to the interaction.
- Players in the information marketplace may be encouraged to promote access to information resources and more fully commercial information distribution in the one sector crucial both to government and the private information companies. This is in the area of information for small- and medium-sized enterprises, and many now believe that this can only be achieved on an international value-added basis through the use of ubiquitous networking. The drive to get things right in this context has so far been ineffective. Countless programmes, funded by Member State governments and the European Commission, have sought to create local resource centres and information dispersal points based on Chambers of Commerce and other entities. By and large these have failed to change anything. A considerable body of opinion now believes that if the value of information, products and services is to become effective within the business models of SMEs then it will happen because the private sector information industry effectively markets to them. And if the private sector information industry is to be re-energised effectively to create the value-added services which SMEs will use then radical easing of access regulations governing government information is urgently required.

Clearly, then, we can conclude that neither the economic opportunities nor the industrial growth requirement have gone away in the last ten years. Indeed, in the age of the information superhighway (a term now only used by politicians and civil servants), this all takes on a desperate relevance. Three case studies drawn from the United Kingdom (once again in the lead, both as an information marketplace and as a regulatory environment under pressure in this regard) demonstrate how the effective dialogue between public and private sectors is now breaking down:

1. In regard to the release of the Register of VAT-Paying Companies in the UK, HM Customs & Excise has, over a ten-year period and in the face of the Government Tradable Information Guidelines, promulgated by another department of the same government, deployed the whole range of possible reasons for obfuscation or delay. When initial claims that the Register was Crown Copyright were rebutted, the Department took refuge in the idea that the information was confidential. When the information industry indicated that it did not seek any confidential data, but simply the complete listing of the names and addresses of companies who currently pay the tax, shorn of details of their registration numbers and the amount of tax that they pay, the Department responded that this data was part of an implied contract between the tax-paying company and the tax collector, and to divulge it would be a breach of that contract. When this argument was demolished, the Department sought refuge in the idea that the data was covered by the Official Secrets legislation, and, indeed, indicated on one memorable occasion that the legal opinion that this was so was itself covered by such legislation. Finally, in 1995, the Department was cornered and forced to relent, and initiated a public consultation on the issue. In April 1996, this consultation came to an end and the Department wrote to participants indicating that 'the core of support for (public access to the VAT Register) came from the finance, credit and information sectors who felt that such access would benefit businesses, particularly those providing business information, assessing creditworthiness, or underwriting potential business risk . . .' Particular issues of concern (there were only 100 responses to this exercise) were privacy and security and 'the additional risk of exposure to unsolicited junk mail or nuisance phone calls.' On these latter grounds it has been decided to maintain the secrecy of the VAT Register. It is one of the beguiling ironies of this UK example that a country which has no central business register, no mandatory registration of trading entities who do not seek limited company status, and therefore no ability to say how many small- and medium-sized enterprises it actually has, feels unable to

release to the information industry data which would improve the accuracy and effectiveness of business databases throughout its economy. And this on the grounds of privacy, security and the avoidance of junk mail in a country with a thriving direct marketing business, excellent business telephone directories, Yellow Pages, and Data Protection legislation which does not extend to corporate entities. As elsewhere in the European Union, it sometimes seems to the information industry as if any argument is good enough to block access to information which can only improve the economic life of the Union as a whole.

2. A further example shows government acting on the other side of the equation. In the early 1990s the Inland Revenue in the United Kingdom, responsible for the collection of all business and personal taxes except for VAT, decided to try to increase earnings from its information resources and in particular from its tax manuals. These documents were internal guidance notes for tax inspectors which advised them on the Department's own view on the interpretation of statute and case law. The tax publishing market in the UK is highly competitive, with some four major groups sponsoring competitive services. For tax accountants and lawyers only fully comprehensive services will do, and if the previously confidential Inland Revenue tax manuals were to become available the information providers felt that they needed to be available in all services. The values upon which publishers were competing in this sector were cross-referencing and commentary, not underlying information. However, in order to get the best price for its product, the Inland Revenue chose competitive tender and awarded an exclusive electronic publishing contract to one of the four competitors. It was then sued and, in a settlement out of court, agreed that a further publisher could have access to the same data. Presumably, if the remaining two competitors wish to gain access, they too will have to bring a legal process (although, hopefully, this reversal will prompt some re-thinking). However, the whole episode, finally resolved in May 1996, illustrates clearly the point made above regarding the increasing willingness of governments to trade information which is essential to key business functions in ways which could lead to the lowering of quality standards in the marketplace as a whole, even if they increase the revenue flow to central government itself.
3. A third example must be the ongoing privatisation of HMSO in the UK. We noted above the creation of data monopolies through privatisation and the problems that this created. In the case of HMSO, Crown Copyright has been specifically removed from that organisation prior to privatisation and a new licensing function will be created within the UK Cabinet Office. So, on the one hand, HMSO will be privatised without the benefit of its unique hold on UK government copyrights, which would seem to diminish its value, while on the other hand, where it does retain control of value-added services which it developed over the years and which UK private-sector interests were unable to develop, it will effectively be privatised with its information monopolies intact. Since presumably the licensing regime will prevent those monopolies from being broken, many believe that this solution provides the worst of both worlds. This solution also does little credit to a generation of adept and diligent publishers at HMSO who have used their skills to create margins which are now worth the effort of privatisation. On the other hand, the privatisation of HMSO in this peculiar manner does not add a new competitive element, or increase the diversity and variety of the UK marketplace.

But one thing clearly signalled by HMSO's repositioning is the determination of many units of government to get their information into shape and onto the Internet so that they can communicate directly with many classes of users. The latest organisation in the UK to take this stand is the Select Committee on IT to the House of Lords, and the first legislation in progress to be treated this way, the Arbitration Bill, has now been published on the Internet. Yet it appears to be the case that governments still want their cake and the ability to eat it. While making this information available in the cause of 'open government', the residual licensing body left behind by the sale of HMSO will undoubtedly continue to seek to sell Crown Copyright. In a recent submission to the European Commission from the General Council of the Bar, the governing body of barristers in the United Kingdom, the barristers noted that 'a crucial issue was whether the continuation of Crown Copyright protection in the law of the land can be justified'. The position in other countries varies but there is some consistency in the overall picture, with the exception of the UK and certain other countries having British-type laws, including India, Singapore and Israel. In the USA there is no state copyright in legislation and the same is understood to be the case in Japan. Some Commonwealth countries have moved, or are moving, away from state copyright protection for their laws; for example, there appears to be no such protection in Pakistan and Australia seems likely to abandon it. In Europe, with the exception of Ireland (where there are indications that copyright in legislation is not enforced), enquiries conducted in recent months have not revealed any country in which the state enjoys copyright protection in legislation. Responses from some 13 out of 15 European Union or EFTA countries indicate that there is no such protection. 'Access to all such information should be unrestricted and it should be available from any publisher who is willing and able to provide the service, with complete freedom of competition. The role of central government, in particular in connection with the "information superhighway" is to take the steps required'.

The irony of this situation in the United Kingdom is that the UK government has a Code of Practice on Access to Government Information and widely advertises its 'open door policy on information'. The Code of Practice on Access to Government Information includes five commitments: to supply facts and analysis with major policy decisions; to open up internal guidelines about departments dealing through the public; to supply reasons with the administrative decisions; to provide information under the Citizen's Charter about public services, what they cost, targets, performance, complaints and redress; and to respond to requests for information. Here, indeed, the chronic schizophrenia of modern governments is fully exposed: on the one hand an Open Door and, on the other

hand, a retention of state copyrights and restricted licensing policy. The information industry at large is now totally confused by these activities. Is there any chance that a French or German database developer, seeking to translate UK legislative materials into the database in his own language for the support and guidance of cross-border traders, would have any hope of understanding the nuances of UK government information policy?

The indignation of the UK barristers is also noteworthy as a symptom of increasing user resistance to government information controls. Barristers will wish to adopt legislative texts of legal instruments into their own information systems and use them for their own reference purposes. They want to buy information services from publishers for the commentary and other value-added services that they provide, and not for the value of the underlying documentation. They believe that government should not restrict their access to that, and it is hard not to agree with them.

In the light of all of these issues, the burden of expectations upon the Green Paper on access to information is very heavy indeed. In early drafts the Commission has indicated its interest in providing more transparency in this area, and the private-sector information industry in Europe has consistently, through the European Information Industry Association, urged the adoption of a directive in this area to replace the ineffective guidelines promulgated over the past ten years. The 19 key principles which information companies in Europe seek to have included in such a directive are attached to this paper as an Appendix. They take the principles cited above and expand them into a matrix of information market measures designed to cover the key issues. While it is as yet unclear as to what stance the Commission will take in the final version of the Green Paper, the fact of the Green Paper itself indicates that rumbling discontent in this area is now widely recognised, and that the Commission believes that this is not a devolved issue of subsidiarity but something upon which important economic interests in Europe rests. Given that the likely drafting and process time required for a directive will be up to five years, it is also clear that the hopes and aspirations of 1986 may have to be postponed a little longer and that we will have to seek a resolution beyond the millennium.

David R Worlock
President of the Council
European Information Industry Association
c/o 26 Rosebery Avenue
London EC1R 4SX
UK
Tel: +44 (0)171 837 3345
Fax: +44 (0)171 837 8901
E-mail: eps@epsdemon.co.uk

References

- Government Tradable Information*, Department of Trade & Industry, London, 1986.
Guidelines on the Synergy Between the Public and Private Sectors in the Information Marketplace, European Commission (DG XIII), Luxembourg, 1989.

Appendix: 19 key principles developed by the EIIA

1. Public organizations should, as far as is practicable and when access is not restricted for the protection of legitimate public or private interests, allow public database collections to be used by private sector companies and exploited by the information industry through electronic information services.
2. Member States should incorporate such a right of access into the conditions of release, use and exploitation of public sector data and information.
3. Member States should regularly review the data and information collected by their departments, and publish the outcome of such reviews in order to further the possibility of the re-use and re-exploitation of government-held information.
4. The principles of licensing and pricing such public data should be made clear by Member States to private sector organizations, and efforts should be made to harmonize these elements across public administrations.
5. The public sector should not place restrictions upon the type of customer or the territories to which newly-created services developed by the private sector from publicly held information sources are sold.
6. Public sector databases shall not normally be sold upon an exclusive basis if exclusivity distorts competition.
7. The public sector should encourage investment by the private sector in the development of information services based on public data.

8. When public sector bodies provide electronic information services directly, such services should not distort competition and should not be created in order to compete with pre-existing private sector services.
9. Where the public sector currently supplies electronic information services, such services should be reviewed, with a view to deciding whether they are more appropriately provided by the public or private sector, and whether greater involvement in the development of such services by the private sector is desirable.
10. Electronic information service providers within the European Community should be treated upon an equal footing and have an equal right of access to the public information held by each Member State.
11. Public sector support for information services should be in line with articles 92 and 93 of the Treaty (on aids granted by the State).
12. Direct or indirect financial support may be provided by the public sector to encourage pre-competitive research and development and the emergence of new market sectors, but should not support the viability of mature services.
13. Public funding may support services which are necessary for the public interest, provided that such services could not be economically viable and could not be provided by private sector interests.
14. There should be encouragement for joint-ventures between public and private sector interests where these are most appropriate for the development of the information marketplace.
15. Conditions governing application of public support to users of European electronic information services should not discriminate against these services on the basis of their European country of origin.
16. Public sector accounting and budgetary procedures and exchange controls should not prevent access by interested public departments to electronic information services throughout the community.
17. The public sector should seek to remove legal and other barriers to the use of public information by the private sector.
18. The public sector should, to the highest extent possible, make use of the discretion given under article 2(4) of the Berne Convention to exempt from copyright all texts covering legislation, public administration, economic and social policy and activity, and norms and standards. Where government-held information does fall within the copyright convention, the public sector ought not to award exclusive rights or reproduction to a single organisation, as this might hinder value enhancement by other commercial sector operators.
19. When public sector information or data is made available for private sector use or exploitation, any pre-existing citizens rights of access to the original information as determined by legislation must be preserved.



Information managers/librarians in the year 2006: prophets, princes or Poohbahs?

Arthur Winzenried

Lilydale Adventist Academy, Australia

Abstract: This paper reports an as yet unfinished Delphi investigation of information management futures. Too often the personal aspect of information provision is overlooked in the light of the dramatic development of technology. However, behind every technological advance there is a provider whose role it is to source, organise and disseminate the information gleaned via that technology.

While the answer to this problem will be different for everyone, some considerations are becoming increasingly important. Each has significance as we move forward. Clearly the Internet is now an important medium and information management needs to include this device. Online data is increasingly providing an economical means of accessing large quantities of data, though not necessarily such as is precisely what our individual clients require.

Perhaps the answer lies in two directions, both needing consideration. On the one hand we need reliable equipment and perhaps a new generation of electronic data technicians to make that equipment and its networking reliable. Secondly we need sound software support that allows integration of data into Internet useable formats, incorporates the Net itself and communicates readily yet is able to be catalogued, refined and maintained in a client-based, easily managed way.

In either case, the need for an information professional is still present. In fact, the need for an intermediary who accesses, assimilates, reformulates and then packages information in a form to suit the end-user is essential if meaningful outcomes are to be achieved. The alternative is chaos.

Keywords: artificial intelligence, information management, Delphi, information literacy

1. Introduction

One problem that besets the modern investigator of the information field is that of definition. What's in a name? someone once asked. But there is an important issue here. Terminology can sometimes obstruct the view and sideline effective consideration of this important issue. In the educational context one often hears of Teacher/Librarians or Media Specialists. In the more open field of business one can occasionally refer to Information Manager or Data Manager in much the same context. For the purposes of this paper, I am perhaps trying to be all things to all people (a very uncomfortable position I believe) and refer to a person whose primary task it is to collect, catalogue and disseminate information effectively from many sources to users with many needs. I will attempt most usually to refer to such a person as Information Manager, more as a means of finding common ground than as a specific statement of vocational nomenclature.

The field of information provision is a changing one. Canadian Teacher/Librarian Lucinda Lockwood put it rather succinctly just recently:

'WHO'S STRESSED!!??? Ha ha:~). Seriously now, it used to be (in the "good ol' days") that once a librarian knew the system (Lib. of Congress or Dewey and Sears) and the process, she had it made. Things were static, and new acquisitions required no further skills to employ them. Now, librarians have to do all of the traditional stuff, and learn new information technology, too! PLUS: she has to do her own trouble shooting; teach the patrons how to use the information (she didn't have to teach patrons how to use a book!); lobby for more money to buy more technology, as well as for traditional, print resources; apply for grants for more money . . .; then there's advocacy, wherein she has to justify her role as a professional in the school. Stress is the word for it!'

For many years her outline, though highly simplistic, was fairly accurate. Life for the information provider was simple and straightforward. The 1990s have changed all that. Today we live in an information rich world. One only has to wander around the main halls of this Meeting venue for an hour or two to be convinced that we live with a virtual information overload. More than two hours there and it becomes necessary to retire to Harrods for some peace, quiet and meditation.

Into this environment, then, the information professional is thrust. It is their responsibility to make some sense

Online Information 96 Proceedings

Page 7

of it, interpret it and then package it in a form which is comprehensible and valuable to the end-user. Further, in these days of financial constraint the end product also needs to be affordable. Nowhere are these constraints more of a consideration than in the field of Education.

Education is that strange animal which produces no visible profit, absorbs more money than anyone reasonably expects it to and which every new incoming government enjoys reviewing as an ideal area for cost cutting. The problem also is that it is that same animal which is the training ground, the preparation workshop for future employees and magnates alike. Their future depends on adequate training, experience and resourcing in our information rich world.

With the rapidly growing cost of technology, its attendant constantly-changing nature (how many versions of Windows have we seen in the past few years?) and the unparalleled growth of Internet resources I fear we are rapidly approaching a repeat of the Great Depression of the 1930s — this time an **Information Depression**. Never, it seems, has the gap between the information rich and the information poor been so great; and it is getting greater all the time.

Into this environment then, the information professionals are thrust.

What is their future? Will they become redundant — rendered so by increasingly simple communications packages and the growing number of online services, tailored for other people's needs?

Will they in fact become puppets dancing to the tune of invisible technologies.

What does the future hold?

I believe there IS a place for future information professionals. I believe its importance is growing with the growth of information, with the widening gap between the haves and the have nots. I believe that role is a changing one, an evolving one but a crucial one. I feel it has a little of the prophet, the priest and the Poohbah in it. I find also that I am not alone in this. There is some considerable agreement among leading professionals, shown by the results of this recent Delphi study.

2. The task

For the present specialist the task can seem enormous. Jamie Mackenzie, in his electronic article, *Libraries of the Future*, summed a modern Library mission up in the following way:

'The mission of the library media program is to ensure that students and staff are effective users of ideas and information.

1. Provide intellectual access to information;
2. Provide physical access to information;
3. Provide learning experiences that encourage users to become discriminating consumers and skilled creators of information;
4. Provide leadership, instruction, and consulting assistance in the use of instructional and information technology;
5. Provide resources and activities that contribute to lifelong learning;
6. Provide a facility that functions as the information centre of the school;
7. Provide resources and learning activities that represent diversity' (Mackenzie, 1995).

The task is large and in the winter of our days can seem beyond us. New forms of data provision can at first glance seem to meet all our dreams, but in the cold light of installing the newest CD-ROM (not compatible with our existing computers) or finding that one elusive piece of information during a fruitless all-night session on the Net, the trees can indeed seem as bare as the few on a Yorkshire Moor in January.

3. The prophet

One of the more important issues facing information personnel today is their ability to predict what is likely to be needed in the future by their clients, and how such information is to be first acquired and later distributed.

In this regard the information manager of the future will have to be something of a prophet. Over the past twelve months I have been 'leading' a Delphi team to investigate information personnel futures. The insights provided by this process have been intriguing. Two aspects have made themselves abundantly clear. On the one hand information professionals of the present are facing more stress than ever before in terms of learning curves and new, additional tasks presented to them. In the face of this, however, they are for the great majority prepared to work, even actively seek, new skills, new technologies and new visions. Their concepts of the future are often quite challenging.

Here are some predictions made by the Delphi team:

- *2 years from now:* High speed broad-band Internet access producing real time movies and live camera action and video-conferencing;
- *4 years from now:* Early forms of so-called 'intelligent agents' are already in use and it is possible to imagine an 'intelligent agents kit' assembled to order by the librarian in order to create a body of knowledge for a specific purpose, and keep it updated. This would involve sampling online-resources at

intervals but would need supervision by an information expert. Such software might not quite be new technology but would support the work of information experts;

- *10 years from now:* virtual reality — virtual trips to inaccessible places with some semblance of reality. Also development of communications globally leading to immediate access and exchange with partners in other classrooms with different experiences and languages. Access to expertise and experience via realistic 'avatars' (Grey (not yet published); Hird 1996; Haycock 1996; Eisenberg (not yet published)).

If any or all of these or any other possibilities are to eventuate, it behoves the information manager of the present and even more importantly of the future to be in the box seat, prepared for the contingency, forewarned and forearmed. Indications are that current information specialists are prepared for change, prepared for the possible learning curve and in some cases already working toward new solutions and technology applications.

4. The prince

Knowledge is power, so they say, and the information managers of the future are certainly going to be responsible for a lot more information than have been their forebears. In this regard they will become information princes, ruling their kingdoms by day and, increasingly in the future, by night also. Information is already becoming a 24 hour-a-day matter (LAA is one of these).

Leadership is required — leadership in collaboration; leadership in the use of new media; knowledge of sources of information and their characteristics (Haycock 1996). Leadership is also required in the area of information selection. It is not enough in view of educational outcomes simply to place a student in front of an Internet terminal and say 'have fun'. Confusion is then the only likely outcome. Outcomes need to be planned in terms of progress along the way, situations and circumstances which will promote the required outcomes, and evaluation of progress at all stages so as to ensure the outcomes do in fact outcome.

This is not easy. It requires considerable skill on the part of the information manager in enabling the information seeker to have access to the appropriate data, to have it in a way that can be comprehended and to have it at a time when it is needed. Technology has enabled more information and made delivery of that information almost immediate. It has not however tailored that information to any particular student's specific needs. This is the important task of the informational professional.

Technology and information need to be thoroughly integrated into curriculum. Franklin Bobbitt (in Jackson 1992, p.7) spoke of curriculum as a 'series of things which children and youth must do and experience by way of developing abilities to do the things well that make up the affairs of adult life; and to be in all respects what adults should be'. The Australian School Library association guide, *Learning for the Future*, suggests that 'resource-based learning is a methodology that allows students to learn from their own confrontation with information resources' (ASLA 1994, p.3). Modern curriculum has taken early views a long way, but the fact remains that today's education relies more and more on integrated learning experiences arising from a multitude of different resources.

Among the problems facing your information prince, though, are the difficulty of controlling his or her subjects. Just when the kingdom is running along smoothly and all the king's men are happily smiling at their computer screens full of information, some laterally thinking, non-intellectually challenged individual will be certain to invent some new piece of software or demand some information not currently available and thus throw the whole realm into a ferment. In our current information world one of these subjects upgrading his software on a more or less annual basis is alone sufficient to keep the realm in something resembling constant turmoil. Your information prince, then, becomes monarch of all he surveys — a constantly changing and challenging anarchy.

5. The Poohbah

One way of challenging this state of anarchy is to become Lord High Executioner — off with their heads! However this does not make the problem go away. In the first instance there is the technology itself making demands often far beyond the specialist's 'usual' training.

'Give me a technician or give me death! Our school is packed to the rafters with computers and related technology. Many teachers are good trouble shooters, but our knowledge only goes so far. That's where a specialist must come in. My library clerk knows less about computers than any other staff member in the school, and the library has the greatest number of different applications! How can this happen?! I love information technology, but glitches such as inadequate technical support will hold us back' (Lockwood 1996).

In the past, information providers trained to access material, enter it into a catalogue and then guide users to that catalogue have predominated in the field. In the more recent past many of these same people have needed to become familiar first with increasingly complex computer catalogue software, then with CD-ROMs, then with networks to integrate the two and finally with the Internet and a host of online services. To add to the troubles some face is the penchant for upgrading software at the rate of new versions almost annually. If you run two programs, upgrade each and put in place new licences for same, install them on a network of 20 computers, fine. But what happens if you run Windows NT, Win95, a library software utilising MS Access and Sequel Server, Office

95 and MS Works for client workstations, Corel Draw 6, EBSCO Magazine databases, Internet access and e-mail as well as a dozen or more CD-ROM titles and 50-plus user oriented educational softwares?

'I see the information specialist as a broker of sorts. With the volume of information readily available increasing rapidly, the individual information consumer will need an increasing amount of assistance to locate, evaluate and synthesise the information he/she needs. If schools shift from a content delivery to a generative curriculum model (as the vast supply of information allows), the information specialist will play a critical role in helping to design and implement the curriculum' (Hird 1996).

The Delphi team highlighted the following skills as important and needing continued attention in view of their importance.

Information delivery 10 years on; important areas:

- (a) More computer and evaluating skills. 4%
- (b) Client support/interaction. 16.6%
- (c) Database/multimedia search/retrieval. 4%
- (d) Information management, esp. teams. 12.5%
- (e) Technology components. 33.3%
- (f) Lobbying and Advocacy. 4%
- (g) Information literacy. 12.5%
- (h) Managing Change. 8.3%
- (i) Internet skills. 4%

One further consideration deserves mention: the often quite considerable gap between administrative expectations and their actual financial contributions. Blowout in information supply, largely due to the advent of the information superhighway, has caused a massive shortfall in data provision at an institutional level. Faced with a superhighway, most institutional drivers are equipped only with Mini Minors! Modern educational establishments not only need heavy-hauling semi-trailers to travel this highway but they also need some means of sifting through all the goodies these trucks can bring home. Too frequently, institutional administrators very well used to the old days continue to provide only the finance for the minis but expect the delivery of a Kenworth!

Success with all these difficulties either leads to something approaching Poohbah status or a prolonged holiday — a specially equipped rest home where they hand out those lovely white canvas tee shirts with the especially long sleeves that do up at the back!

6. Some important considerations

The movement towards a high-tech future in libraries is seen by the Delphi as likely to affect in definite ways the physical plant involved in providing educational information to clients. Already we are seeing two moves of some significance. On the one hand there are a growing number of establishments being built which do not include a formal library or information centre. Rather, information is provided within teaching areas on the basis of being thus nearest and most convenient for users. Information provided in this way is obtained online from sources outside the physical establishment.

Jamie Mackenzie noted this trend, describing it as:

'The worst case scenario is that new technologies and electronic access to information threaten to eliminate both school "libraries" as we have known them and those who have been serving as information tellers' (Mackenzie 1995).

First round results in the Delphi suggested the following list of possible outcomes for 2006:

9. Do there need to be changes in educational structures (physical plant), organisations or educator training to cope with information delivery in 2006?

- (a) Yes — schools without walls. Less as physical school, more as organisation, a combination of services. 18.7%
- (b) Yes — school design, timetable, accessibility, IT facilities, resources, curriculum and delivery. 18.7%
- (c) Yes — amalgamated information/technology resources. 6.25%
- (d) Yes — more small group facilities, project work. 6.25%
- (e) Yes — strong centralised information facilities, well funded. 6.25%
- (f) Yes — different for every place. 6.25%
- (g) Yes — info access to every room. 12.5%
- (h) Yes — a balance between every student access and privacy important. 6.25%
- (i) Yes — emphasis on long distance. 12.5%

Online Information 96 Proceedings

(j) Yes — more emphasis on child centred. 6.25%

By collapsing the more closely related items into one another, the team began to focus more on the child-centred approaches involved in items (b), (d), (h) and (j). However, the values of technology in terms of assisting distance education (and thus the value of some 'libraries without walls') was not lost from view.

The initial response, showing the awareness of the 'libraries without walls' concept, reflected the fears of Mackenzie and raises the issue of control. To what extent is teaching going to be interrupted or adversely effected by access to unrestricted data — information that is not previously selected, tailored and designed for specific needs within a particular institution? In this scenario, where does the role of teacher-librarian exist? As an extra class teacher, as a roving specialist or as a member of a new curriculum entity, a new school department — 'information literacy' perhaps?

The second move being seen currently is that trend towards distance education. While this has many desirable advantages, especially for the isolated students whose geographical location makes any alternative very difficult, there is nonetheless some distinct disadvantage for a student's education where active interchange with teachers and other students does not take place. Advances in technology are allowing more and more intercommunication and there are those (Imagine 3 software developer, John Devine, for example) who are seeing a virtual reality situation where no student physically attends a school and where no teacher does so either. In this futuristic scenario, the implications for interpersonal relationships would need to be totally reconsidered.

What is perhaps needed today more than at any other time of our information history is access to the wealth of data available through a single interface, and allowing a range of search formats from the relatively simple ones needed for our children through to the complex ones required by industrial researchers and businesses.

There is a current trend in Australia to view education as resource-based. Curriculum moves currently taking place, led perhaps by the *Statements and Curriculum Profiles for Australian Schools* and their ramifications, have important consequences for educational outcomes in the future. They also are going to make considerable demands on the type and implementation of data management systems in the next few years. One commentator summed this up in the following way:

'Literacy issues seem also to be very closely linked. The need for students to be not only literate, but information literate be able to search, acquire, organise, interpret and present information — has never been so important as it is now' (McCausland, p.17).

Such emphasis not only on the data but also on the methods of access to the data will become increasingly important as technology continues to impact on curriculum. Workplace pressures demanding more technologically aware graduates will in turn require more active incorporation of technology, especially data access and processing, in the educative process. Integrating this learning of handling skills as well as collection results seems to form an important part of the future as spelled out in the *Profiles*.

I believe similar trends already exist in all major developed countries. Certainly the workplace pressures are very little different from country to country. With this move towards integrated curriculum, information managers are in an ideal position to promote learning experiences by providing the environment and conditions that would favour evaluation of data, its interpretation and its integration into the learning process. The information manager who offers such 'information mediating support to a team of curriculum planners is modeling the kinds information skills we hope students will acquire' (Mackenzie 1995).

7. One solution — the system

One possible solution to the various difficulties is provision of a single interface to the vast galaxy of information currently facing users. Coordinating this on an international basis is probably well beyond our wildest dreams. However, on a local basis this is possible. An integrated multimedia package would need to support the Internet fully and integrate with it. It is with this in mind that LAA has chosen the Imagine 3 software which provides an open-ended package, able to incorporate all forms of current media and organise it within a single search interface — one that is familiar to its users, one that integrates with other commonly used programs and allows seamless operations with them.

It would also need to be supported by sound knowledge and on-site technical help in terms of networking, etc. This could become a joint venture with several organisations, too small to allow budget for one person, sharing a technician.

Possibly this is the key. Costs involved in data provision are rising to unprecedented levels. Unless there is a growth in shared resources, data is quickly going to become too expensive for smaller establishments, whether educational or business. We may see considerably more of the 'haves' and 'have nots' problem, or we may see a new era of shared resources and cooperation.

8. One solution — the role

Early Delphi findings indicated an extension of library/information manager role. Suggestions covered a fair range but the trend was clear.

- (a) More personally client based. 10.5%
- (b) More outside resource connecting. 10.5%
- (c) Environment will change, role still teaching and learning. 5.25%
- (d) Names/titles may change. 5.25%
- (e) Major role in distance learning. 5.25%
- (f) Same as now — collaboration, leadership in new media, knowledge of sources. 21%
- (g) Central. Teaching access and evaluation. 5.25%
- (h) Highly technological — supervising intranets and internets. 10.5%
- (i) Teaching access skills. 21%
- (j) A consultant, a teacher, an information-broker, a cataloguer. 5.25%

Areas of personal service and the teaching of basic literacy skills was considered important. Over subsequent rounds this consideration became more evident as team members moved towards a consensus on a role very much like the present one (item (f)).

There is a future role for information professionals, whether in industry or in education. It is an increasingly important one. As information increases so too does the need for professional management of that information. Accessing, selecting and repackaging for the best use of the end-user is necessary if appropriate outcomes are to be achieved. To simply provide a user with a terminal and say 'happy hunting' is not enough to ensure anything meaningful is achieved. Technology should not control, it should be controlled. It is enormously useful but should not be permitted to become and end in its own right. We may have to become Poohbahs in the age of a new Mikado, we may have to insist on being issued with Kenworths instead of the Minis we are often given, but success is worth the effort that may be needed.

Oh what a tangled web we weave, when first we practice to perceive.

Dr Arthur Winzenried
Information Manager/Librarian
Lilydale Adventist Academy
PMB 1 Lilydale
3140 Victoria
Australia
E-mail: arthurw@ozonline.com.au
Fax: +61 03 9728 6905
Tel: +61 03 9728 2211

References

- ASLA (1994) *Learning for the Future*, ASLA, Melbourne.
- Eisenberg, Mike, Professor, School of Information Studies, Syracuse University, USA, response to the Delphi team inquiry, not yet published.
- Grey, Duncan, Head of Learning Resources, Hinchingsbrooke School, Hinchingsbrooke, UK, response to the Delphi team inquiry, not yet published.
- Haycock, K (1996), Professor and Director, School of Library, University of British Columbia, Canada, response to the Delphi team inquiry, not yet published.
- Hird, A (1996), Director of Partnerships, PEF, USA, response to the Delphi team inquiry, not yet published.
- Jackson, P.W. (Ed.) (1992) *Handbook of Research on Curriculum*, New York
- Lockwood, L (1996), Teacher-Librarian, Maple Ridge, BC Canada, response to the Delphi team inquiry, not yet published.
- Mackenzie, J. (1995) *From Now On: A Monthly Electronic Commentary on Educational Technological Issues*, 5(6).
- McCausland, P. (1995) The implications for school librarians of the curriculum profiles for Australian schools, *Access*, 9(2).



Information professionals as intelligent agents — or when is a knowbot only a robot?

Jessie Hey

Multimedia Research Group and Hartley Library, University of Southampton, UK

Abstract: During the 1980s we saw the role of the information specialist develop with the gradual introduction of end-user searching. The number of online databases was increasing but they were structured and conducive to controlled searching. We now see an explosion in the resources becoming available and being developed by computer scientists, subject specialists, librarians and commercial companies or often by all four together. The challenge for the information professional will be to keep abreast of the most significant developments and to distil the information for a wide range of users. It will be increasingly necessary to pool knowledge with both professional colleagues and clients. The parallel efforts of academics and librarians to develop subject guides to the Web illustrate this well. The computer scientists are currently making dramatic use of computer techniques and computer power to make information more accessible and navigable. Librarians can assist by sharing their knowledge about the efficient organisation of information.

This paper looks at some of the exciting developments in electronic libraries on both sides of the Atlantic and the impact on the information professional. These projects are tackling multimedia, hypermedia, user interfaces and other aspects of the integration of diverse libraries, and showing that methods of research are changing as a result. Information professionals need to see these as an increasingly rich set of resources with a greater variety of techniques for accessing. However, they must continually keep aware of new techniques and sources themselves if they are to contribute to their usefulness. Intelligent agents are beginning to replicate some of the tasks such as current awareness which are familiar to librarians but these can also act as a tool for the librarian (just as driving an automatic simplifies driving a car). The challenge for the researcher and developer and also for the librarian is to make an efficient transport system for the information superhighway; to make it easier to drive different cars and to provide an complementary (and often complimentary) taxi service. Knowledge robots and 3D visualisation are tools to simplify the increasingly complex and diverse information world but pooling skills and resources will become ever more essential.

Keywords: Information professionals, librarian skills, intelligent agents, software agents, knowbots, digital libraries, electronic libraries

1. Introduction

We are now celebrating 30 years of using computers in libraries and information services and are using them as tools in ways we barely dreamed of in the sixties. We expect a library's catalogue to be online and probably accessible from the other side of the world — if we can find the correct route in, around and hopefully out again. We are beginning to use electronic journals and other full-text databases more frequently. We are exploring the potential of databanks and a variety of media. An acoustics journal issue may come with examples of sounds on a CD. Now an encyclopaedia, on CD-ROM, has a video clip showing a space shuttle launch. Personal computers are more powerful than many mainframes used to be and the world is our oyster if we have access to the Internet. The World Wide Web on the Internet has accustomed us all to using hyperlinks as one of the commonest methods for finding our way around this vast distributed multimedia information system. The new seven storey New York Science, Industry and Business library will enable visitors to search more than 125 business and science databases, browse the Web on 100 workstations or bring in their own laptop and use one of 500 seats wired for their machine. During the 1980s we saw the role of the information specialist develop with the gradual introduction of end-user searching. The number of online databases was increasing but they were structured and conducive to controlled searching. The new 'cybrary' in Manhattan symbolises the explosion of resources becoming available and the library profession's initiatives for dealing with it. The new or enhanced information banks are being developed by computer scientists, subject specialists, librarians and commercial companies or often by all four together. They are increasingly incorporating intelligent agents to help users fulfil their information

needs more effectively. The challenge for the information professional will be to keep abreast of the most significant developments and distil the information for a wide range of users. 'To participate more fully, librarians will need the help of technologists to understand the possibilities being created by digital technologies, and technologists will need the help of librarians to appreciate the richness of traditional librarianship and to identify the aspects of it that are most relevant to the continued evolution of libraries' (Ref 1). I will look first at one example of an electronic archive that is changing the way in which research is communicated.

2. Digital library techniques transform scholarly communication

In the high energy physics community, mainframe computers used to be used by librarians (using punched cards and batch processes) to create regular lists of reports (called preprints) to distribute physically around the world. These days the reports themselves are sent electronically by their authors to an entirely automated central depository at Los Alamos National Laboratory in the US (Ref 2). Started by Paul Ginsparg in 1991 to serve a small community of less than 200 physicists, these archives now contain numerous other physics databases and serve over 40,000 users worldwide. The established researcher will go straight to this archive by the most efficient route while a librarian will provide pointers for the student or researcher from another discipline, perhaps by providing links from a subject guide on the library's Web pages. Putting scholarly papers onto the Web will dramatically increase the speed of research communication (although alternative economic models may be needed to sustain the related electronic journal publishing) (Ref 3).

The phenomenal success of the physics e-print archive has motivated Stevan Harnad at the University of Southampton to start an international archive for the cognitive sciences modelled on the American pioneer. It is indicative of the variety of skills that are frequently needed now in the digital information world that the electronic document specialist who will set this up needs a good understanding of a complex set of text processing formats, document conversion, hypertext linking and information retrieval as well as Unix programming tools. This archive will soon join the increasing set of distributed resources to which libraries and their customers will have 24 hour access and which can enhance the productivity of research.

3. Evolving information services and the demand for types of information skills

Discovery tools are important in a distributed environment, although to some extent their significance is decreasing owing to the growth of digital libraries. Digital libraries such as those above are usually established by individuals or bodies that care enough about their contents to build, maintain and in some cases describe them. An example of a body developing a library with an associated 'electronic community' is the Interest Group on Pure and Applied Logic, an international association of almost 1000 logicians from almost all the countries of the world (Ref 4). Thus, for many users, one option for reducing the task of network discovery is to find the best digital library and then search within it.

Paul Ginsparg suggests that the electronic medium naturally facilitates large scale disintermediation (although his own community uses what to some people might seem quite a high level of information handling skills to transmit and retrieve their reports) but he admits that there remains a need for organisation of *intellectual* value-added services. Current concerns for his archive include not only the continued development of a robust global mirroring system but also a better means of handling meta-level indexing information. In helping tackle indexing problems such as these, the library and information science communities could surely provide a constructive contribution. 'It would be remarkable if centuries of ostensibly relevant experience will find little applicability in the network context' (Ref 5). The skills of the archivist may also be crucial for addressing the issues of the long term preservation of electronic resources. There is indeed optimism that the demand for information professionals, in the broadest sense, will continue to expand in our information intensive society. Michael Hill, Chair of the Aslib conference on the Future Information Professional in May this year, stressed in the invitation to participate in the meeting that 'The information society by definition opens up enormous numbers of opportunities for the information professional. What he/she needs is foresight, imagination and the courage to go for them.' Nick Moore (Ref 6) suggests that three complementary groups will emerge: creators, communicators and consolidators. We may need to reflect more carefully in which area our skills lie or more consciously add new ones like design skills as part of our career development. We will certainly need to master the fast changing jargon of Internet and multimedia database retrieval in order to teach some of our customers these information skills. Digital literacy means being at home in an expanding but shifting mixture of words, images and sounds (Ref 7).

In a related approach to looking at the role of the intermediary, Brewer *et al.* (Ref 8) look at the value-added services that need to be provided in the emerging digital library. They then divide them up into search services, classification, information filtering, translation services (for formats as well as languages) and publishing. While the digital world will blur distinctions between publishers and libraries they argue that digital libraries will bring many new forms of intermediaries, some rooted in the services that libraries and publishers have always provided and others which respond to entirely new intermediation needs. Future libraries will in any case incorporate both digital and traditional resources. The proactive nature of today's library services will need to be continued even

Online Information 96 Proceedings

Page 14

if some of the intermediary's role may also be to develop software agents to simplify 'robot librarian' tasks in the digital environment.

The working party of the American Coalition for Networked Information (CNI) on networked information discovery and retrieval has endorsed the view that the active, continual involvement of a human will be still be necessary to control the process and to make complex decisions involving relatively unstructured information (Ref 9). Therefore, although much research effort in electronic libraries is being expended in automatic and semiautomatic semantic indexing there will still also be an active role for the librarian or subject specialist. Abstracting and indexing in particular will entail some human input until the day all data is highly structured and complies with clear data interchange standards, and the tools become truly intelligent. In the UK pilot projects in the Electronic Libraries Programme (Ref 10) such as EEVL in Engineering and OMNI in medicine, librarians are developing specialised resources which provide the metadata (the information about information) currently rarely available for resources on the Web. At the same time they are adding value by sifting for quality sources with a UK relevance which are also harder to find in a world with an initially American bias. These subject based services will provide effective pointers for librarians who are teaching students or researchers new to a given field how to use the Internet most profitably in their research projects.

4. Wide ranging progress is taking place in electronic libraries

In the UK, the ELVIRA series of conferences on Electronic Library and Visual Information Research, which has now reached its third year, samples the progression of research in the many interdisciplinary fields of digital library research. These range from fundamental infrastructure issues such as turning the World Wide Web into a library (Ref 11) to the state of the art of automatic image content retrieval (Ref 12), and practical examples such as creating a digital medieval manuscript (Ref 13) or publishing a scholarly journal on the Web (Ref 14). We are now at the stage of having some experience of setting up (Ref 15) and using tailored electronic libraries. Evaluation of how effective they are is beginning to take place, for example, in the Elinor library (Ref 16) at De Montfort University and this is an aspect of digital library work which will profit from more research. The information specialist is well placed to help look at the real needs of the user while the computer scientist creates new methods (Ref 17) of helping the librarian visualise where the end-user is having difficulty.

5. Visualisation in digital databases

Visualisation is increasingly a key area of emphasis in research with interesting experiments in 2D and 3D mapping. Some examples will serve to illustrate this. The Hyper-G hypermedia information system for the Web (Ref 18), initiated at Graz University of Technology in Austria, is developing a 3D information landscape (which can be viewed on a powerful Unix machine) to supplement its 2D local maps as an aid to navigation in hyperspace. At Delft University of Technology in the Netherlands, researchers are using techniques at the confluence of geographical information systems and multimedia to develop a prototype of the hypermap concept of georeferenced multimedia for their campus information system (Ref 19). On the East coast of the US a digital library, also based on a spatial metaphor, is being developed in the State of Florida — metadata on collections is divided into information zones on a map of Florida (Ref 20). On the West coast the Alexandria Digital Library (Ref 21), which is co-ordinated by the University of California, Santa Barbara, will provide online public access to maps, photos and other information referenced in geographical terms. It will help make available data currently found only at major research libraries. The user interface includes a map browser and the search is done by zooming and panning a base map. The spatial indexing will play an increasingly important role as the catalogue grows.

Another interesting experimental interface is being developed for a very different set of spatial sequenced digital data — the National Library of Medicine's Visible Human Project which is in the process of creating a large digital library of anatomical images of the human body. In the Visible Human Explorer user interface (available from <http://www.nlm.nih.gov/>) the library contents are indexed by metadata consisting of automatically generated miniature visuals. By dragging a slider over an overview of the body users select the visual previews from which they can straightaway download the desired images. Human factors research will be needed to determine usability measures of differences between 2D representations such as this Visible Human Explorer interface (Ref 22) and potential 3D representations. However, as computer hardware increases its power, 3D representation techniques are becoming more practical and attractive.

6. Large scale test-beds forge ahead in US digital libraries

In the US the government has made digital libraries the flagship research effort for the National Information Infrastructure. The Digital Library Initiative (DLI) (Ref 23), sponsored by the National Science Foundation, the Advanced Research Projects Agency (ARPA) and NASA aims dramatically to advance the means to collect, store and organise information in digital forms, and make it available for searching, retrieval and processing via communications networks. The DLI projects, such as the Alexandria Project cited above, give a good indication of the current research into large-scale digital libraries. Six projects run by university-led consortia are funded

from September 1994 to August 1998 and all are building testbeds with large collections to relate their fundamental research questions into building large-scale electronic libraries.

7. An established digital library begins to mould research techniques

In the humanities we already have a digital library which has been developing for long enough to give us some findings on the impact of putting a critical mass of materials in digital form. The Perseus Project (Ref 24) is an on-going digital library on ancient Greek culture that has been under development since 1987. Experience has shown that it is almost impossible to predict with accuracy what electronic tools will and will not prove valuable. Services that initially seemed promising have at times been less useful while functions which had attracted little or no thought have emerged as dynamic new tools and even catalysts for the transformation of practice. Does this then indicate that evaluation of new tools, such as Web searching tools (Ref 25) and intelligent agents, is thus a constructive area of research for information professionals to be involved in? What is striking in the Perseus Project is that translating from print into electronic form has undercut the distinction between student and research tool. Analysis, although still at an early stage, is tending to show that students looking up the meaning of a word use the research lexicon and only 20% of the time choose to switch to the briefer entry in the student lexicon. In their paper forms the research lexicon was too dense and complex for use by any but professionals while the student lexicon had previously languished untended for more than a century. With the possibility of incorporating thousands of images cheaply text is also beginning to share centre stage with other electronic documents in this particularly visual subject and, with the development of languages like Virtual Reality Modelling Language (VRML), 3D walk-throughs could become a natural area of development for this database.

8. Virtual reality provides visions of future information worlds

There are in fact many images of the cybrary and the cybrarian of the future which help us envisage some of the potential of the emerging technologies for electronic libraries and distributed multimedia information systems. The DigiCom company's virtual reality databank in the film of Michael Crichton's bestselling thriller *Disclosure* (Ref 26) is vividly portrayed by modern film techniques as a 3D library. In a multimedia version of my paper I would like to be able to link the reader directly to the appropriate video sequence but it is perhaps indicative that the DigiCom library becomes a black and white world when there is too much data to move around. Current research into the real-time display of continuous media data such as video and audio (Video-On-Demand or VOD) is focusing on such issues as improving the transmission of media traffic over high-speed networks (Ref 27) but the technologies and also the copyright mechanisms are not yet quite mature enough to satisfy all our wants. Neal Stephenson's science fiction novel *Snow Crash* (Ref 28) introduces a wonderful piece of software — the Librarian daemon, who looks like a pleasant, fiftyish, silver-haired, bearded man with bright blue eyes. He does all the investigative work of a robot librarian: the only thing he can't do is think. The virtual world, called the Metaverse, is also populated with avatars — the surrogate graphic characters which have become familiar in online chat systems such as CompuServe's Worlds Away. Dedicated Worlds Away users can even upload their own images to be used as their avatar's head.

Another visionary scenario is Martin Halbert's *Knowbot Explorations in Similarity Space: Library Conversations in the 21st Century* (Ref 29). These knowbots, like the Librarian daemon above, are programs which collect information from network databases and organise it for their owners. Textures are used to convey information about the database regions in the 3D datascape. Barren fields, for example, represent very new records in the database that have not been analysed for content, type of publication etc. They have only been scanned and indexed for basic similarity measures to place them in the database. Mountainous regions represent clusters of relevant documents. Buildings represent hypertext links so that a heavily cited journal in an electronic journal database looks like a metropolis. The knowbots are based in a 'knowbot corral' and are represented by animated holographic icons. In the corral are clusters of knowbots, sub-knowbots and metabots. The metabots are knowbots that can be programmed with customers' preferences. They can automatically suggest and manage the creation of new knowbots when they spot a relevant new topic. They can also optimise the current stable of robots. This futuristic technology portrayed here is still in its infancy but it is now beginning to be possible for lay people to build their own knowbots or software agents. Books (Refs 30, 31) can now provide you with enough information about bots and the software (such as Tcl, Java or Telescript) necessary to create them, to customise your very own personal digital assistant or research assistant.

9. Knowbots and software agents

Intelligent software agents are often popularly referred to as knowbots but there are numerous more specialised definitions around, like the sinister sounding:

knowbot: an independent, self-acting computer program that seeks for information on behalf of a user, possibly

replicating itself on other hosts on the network. As the knowbot performs its task, it sends reports back to the user, and self-destructs when it completes its task (Bryan Pfaffenberger — *Internet in Plain English*).

The power of knowbots on the Net is increasing but the early examples on the Web frequently look like basic information retrieval techniques translated into a world that was previously lacking them. When programmed to repeat their searches they may be an upgrade of the familiar Selective Dissemination of Information (SDI) of traditional online databases — updated with a more friendly interface and often more sophisticated searching techniques. The computer scientists have reinvented current awareness profiles for the masses and called it information filtering. An interesting example is the Mercury Center's Newshound (<http://www.sjmercury.com/hound/>) which automatically searches articles from a wide range of newspapers and wire services, as well as classified ads from the *San Jose Mercury News*, using a profile created by you and sends any relevant documents directly to your electronic mailbox. Knowbots on the Net are knowledge acquisition robots — autonomous programs that can search for and act upon information found on the Internet. Knowbots can sit on a single node and probe the Net, or can wander from server to server (as a worm). Examples of currently existing knowbots are Archie servers, some directory servers, the Web meta-indexes and the World Wide Web wanderer. One can now draw a distinction between first generation and second generation knowbots. A first generation knowbot is a tool that traverses the Web and retrieves raw data with no post-processing. A second generation knowbot is a more complex tool that uses some kind of knowledge analysis system, like an expert system, to refine the information extracted from the Internet. While the Newshound shows minimal intelligence, other systems such as NewsWeeder (Ref 32) contain agents which exhibit more intelligence by learning users' interests. WiseWire, the commercial version of NewsWeeder, will be a Web-based service that uses neural net filtering technology to learn a user's interests automatically. As you rate information, WiseWire learns your responses. At the same time all new content is filtered by users with similar information preferences. This collaborative filtering is added to provide a more efficient personalised online information service which they call a Personal Surfing Assistant.

Knowbots, softbots, taskbots, interface agents, personal agents and network agents are just some of the new terms used by various researchers to describe their agents. The idea of employing agents in the interface to delegate certain computer-based tasks was introduced by visionaries such as Nicholas Negroponte and Alan Kay and over the past few decades numerous researchers such as Marvin Minsky and Patti Maes have studied problems that demonstrate some type of agent behaviour. Agent research draws on the results from several disciplines and there is continuing debate within the community on ways of classifying and defining agents (Ref 33). At a simple level, agents can be considered personal software assistants with authority delegated from their users. Patti Maes at MIT's Media Lab, for example, has been working to create agents which reduce both work and information overload for computer users (Ref 34).

There had been a long tradition of investigation into the use of agent intermediaries in the information world before the advent of ubiquitous and powerful PCs with easy access to the Internet which has helped fuel the current enthusiasm. For example, twenty years ago the Rand Intelligent Terminal Agent (RITA) (Ref 35) provided a network advisory system which proved a good testbed for the demonstration and evaluation of intelligent agent capabilities. Software intermediaries were developed within the field of expert systems — computer-based systems that use knowledge and reasoning techniques to solve problems that would normally require human expertise, usually within a relatively narrow domain. That inexperienced users could operate retrieval systems at all was demonstrated by experiments with the CONIT system from MIT. But whereas CONIT taught users a command language and suggested search techniques so that the users then had to decide for themselves how to apply these techniques, their next experimental intermediary system EXPERT (Ref 36) aimed to help inexperienced users search bibliographic retrieval systems using a menu-selection/ fill-in-the-blanks mode to aid search strategy formulation and reformulation. In the UK, Pollitt was one of the first to use artificial intelligence techniques for a database access system. His menu-based system CANSEARCH (Ref 37), using a touch sensitive screen, provided access for doctors to cancer therapy literature in the Medline database using Medical Subject Headings (MeSH) to guide the search process. Another system, TOME.SEARCHER (Ref 38), commercially available from Tome Associates, which was first used to search INSPEC's scientific databases has been extended to provide an intelligent front end to a wide range of databases.

More recent work is often pictured in a more modular way with an emphasis on multi-agent systems composed of a set of intelligent entities called agents. These agents have their own capabilities and then they communicate with each other and cooperate in order to achieve a global goal.

10. Intelligent agents in practice

One useful framework for understanding current work on agents divides them into three types of agent which perform a variety of specialised tasks:

- interface agents
- task or mediator agents
- information agents

Interface agents interact with the user, receiving user specifications and delivering results. Task agents have knowledge of the task domain and support decision making by formulating problem solving plans. They carry out

these plans by querying and exchanging information with other software agents. *Information agents* provide intelligent access to heterogeneous collections of information sources.

The University of Michigan Digital Library project (Ref 39) is building its infrastructure in this way with three separate groups of agents. It also provides a subclass of mediator agents, called facilitators, expressly to team up with other agents to accomplish more ambitious tasks. The 'robot corral' is beginning to be populated. In this case, a shoal might be a more appropriate metaphor since the remora agent is an example of one of the mediator agents in the Michigan library. The remora is a kind of fish which attaches itself to sharks and the remora agents attach themselves to information agents for the purpose of detecting events. Users specify events of interest and receive notification when such events, like new items appearing in the collection, occur. This current awareness service is just one of the value-added services that the library can provide. The agent architecture lets them develop specialised capabilities and add them as needed. For example, through new interface agents they can customise interfaces to new user classes such as high school students.

The ability to perform reasoning is one of the key aspects of intelligence that distinguishes intelligent agents from more robot-like agents. Reasoning, in turn, is dependent on the ability to demonstrate human like traits like emotions, beliefs and intentions. In order for us to feel more confident about the actions and level of abilities of our assistants, some agents are represented by agent icons not unlike the fictional examples previously referred to. One of the first agents with wide recognition, both in the press and with a large number of users, was the BargainFinder Agent (<http://bf.cstar.ac.com/bf/>) from Andersen Consulting, which assists with comparison shopping on the Internet. It uses a miner with his lamp to show progress on a query. MIT has produced icons in which the faces are deliberately simplified to avoid presenting an impression of more intelligence than the agent actually possesses. A 'smart' image could lead one to expect all the services of a real human professional advisor. However, many more information aware organisations are now in a position to evaluate their use of prototype agents and GlaxoWellcome (Ref 40), for example, have found that their 'dumb' agents have proved extremely useful in helping their knowledge workers navigate their heterogeneous information systems. They therefore endorse the view that the first innovations will be in simple time saving but will pave the way for more sophisticated agents and more profound benefits. In the meantime they will pursue parallel initiatives on improving the consistency of their information infrastructure to simplify the problems. Coming back full circle to the very first example in this paper of the Physics community, there is also another successful database project in Germany — a Network Publication System for the European Physical Society (Ref 41) which will allow remote users to order a document from several document databases for local printing-on-demand. Early indications here also show that agent technology can be useful in serving the needs of specific user groups.

11. Conclusion

No one individual or profession has all the skills now needed to create an information infrastructure for a community of users. We have seen many variants in the electronic library and other distributed multimedia information systems projects with partnerships between computer scientists, librarians, teaching faculty, trainers, publishers and media designers. The advent of electronic publishing and particularly the Web has given everyone the potential to be a publisher. However, the opportunities for using multimedia give options for sophisticated presentation that will increasingly put a premium on artistic and design skills as part of the portfolio of skills desired. Hopefully the end result will be resources that the user will require fewer skills to negotiate and manipulate, even given the increasing richness and variety of the material. Knowbots and other agent technologies, whether of the smart or dumb kind, look set to ease but not take over all of the tasks of the end-user. Agent research, in fact, gives a real opportunity to put research results directly into the hands of the end-user. After over 20 years of using online databases the skills and knowledge base needed by both the user and the intermediary in this increasingly distributed multimedia environment will continue to change and keep us on our toes.

Acknowledgements

The author would like to thank Professor Wendy Hall, Hugh Davis, Les Carr and other members of the Multimedia Research group who continually expand my own vision of how computer science techniques will help us make the most of our information rich society. My colleague Jonathan Dale (with his timely M.Phil thesis on a Mobile Agent Architecture to Support Distributed Resource Information Management) helped clarify the key issues in the dynamic field of intelligent agent research and Professor Cherri Pancake of Oregon State University provided encouragement with her user-centred approach to high performance computing.

Jessie Hey
Multimedia Research Group
Electronics and Computer Science Department and Hartley Library
University of Southampton
Southampton
SO17 1BJ
UK
Tel: +44(0)1703 593256

Online Information 96 Proceedings

Page 18

References

- [1] Levy, D.M. and C.C. Marshall (1995) Going digital: a look at assumptions underlying digital libraries, *Communications of the ACM*, **38**(4), 77–84. This is a special issue on Digital Libraries.
- [2] Ginsparg, P. (1994) First steps towards electronic research communication, *Computers in Physics*, **8**(4), 390–6.
- [3] Harnad, S. and J.M.N. Hey (1995) Esoteric knowledge: the scholar and scholarly publishing on the Net, in L. Dempsey, D. Law and I. Mowlat (Eds.) *Networking and the Future of Libraries 2: Managing the Intellectual Record*, Proceedings of an international conference, Bath, 19–21 April 1995. Library Association Publishing, London, pp. 110–116.
- [4] Gabbay, D. and H.J. Ohlbach (1996) Electronic support for scholarly communication: developing an electronic community, in *Networked Information in an International Context*, a conference organised by UKOLN, Heathrow, 9–10 February 1996. British Library R&D Report 6250, 95–98 and the IGPL Web pages.
- [5] Ginsparg, P. (1996) Winners and losers in the global research village, invited contribution for conference held at UNESCO HQ, Paris 19–23 February 1996 during session 'Scientist's view of electronic publishing and issues raised', <http://xxx.lanl.gov/blurb/pg96unesco.html>.
- [6] Moore, N. (1996) Creators, communicators and consolidators: the new information professional, *Managing Information*, **3**(6), 24–25.
- [7] Lanham, R. (1995) Digital literacy, *Scientific American*, **273**(3), 160–161. Issue on Key Technologies for the 21st Century.
- [8] Brewer, A., W. Ding, K. Hahn and A. Komlodi (1996) The role of intermediary services in emerging digital libraries, in E.A. Fox and G. Marchionini (Eds.) *Digital Libraries 1996*, Proceedings of the 1st ACM international conference on digital libraries, Bethesda, Maryland, 20–23 March 1996. ACM, New York, pp. 29–35.
- [9] Lynch, C. (1996) A summary of the findings of CNI's NIDR research initiative, in *Networked Information in an International Context*, a conference organised by UKOLN, Heathrow, 9–10 February 1996. British Library R&D Report 6250, 89–93 and from <http://www.cni.org/>.
- [10] Details of the UK Electronic Libraries Programme (eLib) projects can be found on the UK Office for Library and Information Networking (UKOLN) server at <http://ukoln.bath.ac.uk/elib/> and updates on progress can be seen in the electronic journal ARIADNE at <http://ukoln.bath.ac.uk/ariadne/>, or its paper equivalent.
- [11] Carr, L.A., H.C. Davis, W. Hall and J.M.N. Hey (1996) Using the World Wide Web as an electronic library, in *ELVI-RA*, Proceedings of the 3rd International Conference on Electronic Library and Visual Information Research, Milton Keynes, 30 April–2 May 1996 (to be published).
- [12] Eakins, J. (1996) Automatic image content retrieval: are we getting anywhere? (ibid.).
- [13] Arnott, M., I. Beavan and J. Milne (1996) The online bestiary project (ibid.).
- [14] Whalley, B., S. Munroe, S. Landy, S. Trew and J. MacNeil (1996) Publishing a scholarly journal on the World Wide Web (ibid.).
- [15] Sunahara, H. et al. (1996) A digital video library system at the Mandala library (ibid.).
- [16] Davies, C. (1996) Student conceptions and use of an electronic library: lessons for design (ibid.).
- [17] Twidale, M.B. and D.M. Nichols (1996) Collaborative browsing and visualisation of the search process (ibid.).
- [18] Flohr, U. (1995) Hyper-G organizes the Web, *BYTE*, **20**(11), 59–64. For further information on Hyper-G see: Maurer, H. (1996) *Hyper-G: the Next Generation Web Solution*, Addison Wesley or <http://hyperg.icm.tu-graz.ac.at/hgbook>.
- [19] Kraak, M.-J. (1996) Integrating multimedia in geographical information systems, *IEEE Multimedia*, **3**(2) 59–65.
- [20] Orendorf, J. and C. Kacmar (1996) A spatial approach to organizing and locating digital libraries and their content, in E.A. Fox and G. Marchionini (Eds.) *Digital Libraries 1996*, Proceedings of the 1st ACM international conference on digital libraries, Bethesda, Maryland, 20–23 March 1996. ACM, New York, pp. 83–89.
- [21] Smith, T.R. (1996) A digital library for geographically referenced materials, *Computer*, **29**(5), 54–60.
- [22] North, C., B. Schneiderman and C. Plaisant (1996) User controlled overviews of an image library: a case study of the Visible Human, in E.A. Fox and G. Marchionini (Eds.) *Digital Libraries 1996*, Proceedings of the 1st ACM international conference on digital libraries, Bethesda, Maryland, 20–23 March 1996. ACM, New York, pp. 74–82.
- [23] Schatz, B. and H. Chen (1996) (Guest Eds.) *Computer*, **29**(5). Issue on the Digital Library Initiative. UKOLN also provides a mirror Web site for US Digital Library projects.
- [24] Crane, G. (1996) Building a digital library: the Perseus Project as a case study in the humanities, in E.A. Fox and G. Marchionini (Eds.) *Digital Libraries 1996*, Proceedings of the 1st ACM international conference on digital libraries, Bethesda, Maryland, 20–23 March 1996. ACM, New York, pp. 3–10.
- [25] Winship, I.R. (1996) World Wide Web searching tools — an evaluation, *VINE*, **99**, 49–54.
- [26] Crichton, M. (1994) *Disclosure*, Arrow, London

- [27] Fox, G.C. and J.M. Del Rosario (1996) Constant bit rate network transmission of variable bit rate continuous media in Video-On-Demand servers, *Multimedia Tools and Applications*, 2(3), 215-232.
- [28] Stephenson, N. (1993) *Snow Crash*, ROC, Penguin Books Ltd, London.
- [29] Halbert, M. *Knowbot Explorations in Similarity Space: Library Conversations in the 21st Century*, <http://ls.rice.edu/~halbert/Pubs/ital.html>.
- [30] Williams, J. (1996) *Bots and Other Internet Beasties*, Sams.net, Indianapolis
- [31] Cheong, F-C (1996) *Internet Agents: Spiders, Wanderers, Brokers and Bots*, New Riders, Indianapolis.
- [32] Lang, K. (1995) NewsWeeder: Learning to Filter Netnews, in *Proceedings of the Twelfth International Machine Learning Conference*, Morgan-Kaufmann Publishers, San Francisco, pp. 331-339
- [33] Wooldridge, M. and N.R. Jennings (1995) Intelligent agents: theory and practice, *Knowledge Engineering Review*, 10(2), 115-152.
- [34] Maes, P. (1994) Agents that reduce work and information overload, *Communications of the ACM*, 37(7), 30-40. Special issue on Intelligent Agents.
- [35] Anderson, R.H. and J.J. Gillogly (1976) The Rand Intelligent Terminal Agent (RITA) as a network access aid, in *AFIPS Conference Proceedings*, 45, New York, 7-10 June 1976. AFIPS Press, Montvale, New Jersey, pp. 501-509.
- [36] Marcus, R.S. (1981) An automated expert assistant for information retrieval, in *The Information Community: an Alliance for Progress*, Proceedings of the 44th ASIS Annual Meeting, volume 18, Washington, DC, 25-30 October 1981. Knowledge Industry Publications, Inc., White Plains, New York, pp. 270-273.
- [37] Pollitt, A.S. (1987) CANSEARCH: an expert systems approach to document retrieval, *Information Processing and Management*, 23(2), 119-138.
- [38] Ford, N. (1991) *Expert Systems and Artificial Intelligence: An Information Manager's Guide*, Library Association Publishing, London, pp. 136-144.
- [39] Atkins, D.E. et al. (1996) Towards inquiry-based education through interacting software agents, *Computer*, 29(5), 54-60.
- [40] Abed, S. and R. James (1996) Agents at work — pragmatics of applying agent technology, in *PAAM 96*, Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, London, 22-24 April 1996. The Practical Application Company, Blackpool, pp. 1-9.
- [41] Borghoff, U.M., R. Pareschi, H. Karch, M. Nohmeier and J.H. Schlichter (1996) Constraint-based information gathering for a network publication system (ibid.), pp. 45-59



Library users: how they adapt to changing roles

Helis Miido

International Agency for Research on Cancer, France

Abstract: The library has undergone drastic changes in the last few years: from centralisation to decentralisation; from practitioner to teacher; from passive to interactive information provider, knowledgeable in automation and new technology as it affects not only the manner in which the library performs its functions but, more importantly, the library user. Historical library tasks, e.g. database searching, are increasingly performed by library users, forcing both the librarian and the user to assume at times dichotomous roles of teacher and student (the scientist grudgingly becoming the unenviable student in uses of new technology as the librarian frantically reads the next chapter in the use of Macintosh Powerbooks). Modern librarians install new software and guide organisations in multimedia applications (that is, if the librarian wishes to retain credibility with the clientele). Librarians thus have to be cognisant of the human factor, and how the library users adapt to their changing role as information providers. The experience of one medical library as it implements Windows NT as the library server, running applications such as KR ScienceBase and OPAC, and its effect on end-users is discussed in terms applying new retrieval methods, sources and dissemination techniques, and improving efficiency and cost effectiveness.

Keywords: Library automation, library services, user-computer interface

1. Introduction

The library user has been unexpectedly thrust into the unenviable realm: the lofty heights of experimentation. Like the scientist who runs tests to demonstrate known truths or validate hypotheses, the library user has been studied, debated, criticised and heralded as the technological new-age prototype, able to leap tall buildings in a single bound by simply surfing the net. The inevitability of direct end-user access to information products and services, and the concomitant transfer of said services from the librarian to the end-user, has been affected, like comedy and tragedy, by high expectations, tempered with apprehensions of misuse.

Is there a future for the end-user in online bibliographic searching? we (collectively) asked in 1981 (Ref 1). Is patron behaviour positive or negative? we postulated in 1982 (Ref 2). Is the librarian's credibility affected? we inquired as early as 1984 (Ref 3). Do university undergraduates perform manual literature searches in a costly and more time effective manner than computerised searches? (Ref 4). Does direct online searching increase flexibility, independence and cost effectiveness or decrease them? (Refs 5, 6) Do librarians as intermediaries to online searching really increase speed and currency of the information, reduce overheads, control costs and maximise resources? (Refs 7, 8).

End-users have been integrated into instruction programs (Refs 9 - 11), included in pilot projects (Ref 12) and charged for their services (Ref 13). Have we said all there is to say about the human factor of disintermediation (i.e. the process whereby information providers bypass traditional librarians and offer their services directly to the end-user)?

Technology is advancing at an increasing rate. There is no limit to what end-users can do. However, when faced with this deluge of opportunities (and at times dilemmas over optimum utilisation of time), how does the end-user react? Is it enough to give a one-day course in bibliographic searching or is the process of disintermediation more subtle? Have all the problems, benefits and improvements been addressed, especially as they affect ease of adaptation?

The International Agency for Research on Cancer (IARC) provides an ideal environment to study end-user adaptation to new services from a human perspective. The library is small, with holdings of approximately 4000 books, 11,000 bound journals and 200 on-going journal subscriptions. The librarian and two staff interact closely with approximately 200 library users, 50 of whom are temporary, for example visiting scientists or trainees, periodically replaced as new awards or training grants are announced. Research is conducted on a number of topics including epidemiology, genetics and virology. End-user needs at IARC are uniform in the types of information required and modes of delivery, and thus easier to monitor. Scientists work on PC and Macintosh applications, enriching the problem of accessing library services. The IARC Library used this diverse composition of equipment and library requirements to explore how the end-user adapts to changing roles, and the effect on the

role of the librarian, while converting from a fully manual to a fully automated library.

Today, after automation and disintermediation, the library educates new users in available library services, including PC and Macintosh applications for Medline, Current Contents, KR ScienceBase, Reference Manager, Netscape, FileMaker Pro 3.0 databases, MS Office products and specialised services available on a local area network. However, the road to success was not easy. This story is based on 'Necessity, the mother of invention' (George Farquhar), or rather, 'Progress . . . is not an accident but a necessity' (*Social Statics*). If 'education has for its object the formation of character' (*ibid.*) then this experiment was a resounding success!

2. Initial library automation

2.1. Interlibrary loans

IARC staff publish over 200 papers each year. Obtaining reprints not available in the library has always been of paramount importance. Since the mid 1980s library staff have identified new references through online searches, primarily of Medline/Cancerlit on Dialog. The resulting reprints were not always available in the IARC Library. Requests for interlibrary loans were manually processed, i.e. the library user filled out a request form and library staff either typed the request onto an interlibrary loan system, such as one provided by Dialog, or tried to obtain the item from other local libraries, usually by calling on the telephone. Record keeping — for example whether the requested article was received, length of time to receive requests and so forth — was non-existent. The library carried the cost of this service. As information flourished and the number of published papers increased over time, the number of requested reprints increased as well. The library budget deficit, sometimes amounting to \$10,000, was made up from the travel budget and from equipment and supplies line items. Confidence in the library provision of interlibrary loans was at an all-time low, and budgetary over-runs were rampant. While the number of interlibrary loan requests increased, other library tasks increased as well. Needless to say, the number of library staff remained the same.

'Necessity, the mother of invention.' The library had to increase efficiency and performance. The most obvious change was to eliminate the multiple handling of interlibrary loans. The same reference was re-typed multiple times, by the requester, by library staff into reference-fulfilment systems such as Dialog and for follow-up. The same reference was typed by the secretary for inclusion in bibliographies. Statistics were manually compiled and record keeping almost nil. To complicate matters, administration clung tenaciously to historical tenets. The fear of abstracts replacing full text and potentially corrupting the pure scientific ethic limited automation to basically statistical analysis and wordprocessing using the centralised VAX WPS. Believing the need for efficiency was greater than the fear of insulting the scientific ethic, the library obtained a free UNESCO mainframe VMS database software called CDS/ISIS and established a database for ordering reprints. Database features (described more fully elsewhere (Ref 14)) included a simple Menu for library users for data entry, and a more complex Menu for library staff.

Thus, in 1989, a memo to all IARC staff announced the implementation of an automated interlibrary loan system and offered training in its use. The library user was not given a choice; as of a certain date, manual interlibrary request forms were no longer accepted by the library. A more detailed handout describing the database and data entry procedures was prepared and distributed to all potential users, and brief training sessions proposed. End-users were requested to use the same dumb VAX terminals used by secretaries for wordprocessing. An additional terminal was made available in the library.

The success of this interlibrary loan database was statistically confirmed: 652 reprints ordered in 1987, 1928 in 1991 and 3520 in 1992. The numbers fell to 2202 in 1993 due to the installation of a separate database for ordering reprints for the Agency monographs programme.

How did the end-user adapt? The younger visiting scientists/trainees who had already experienced automation in their home universities or institutes required minimal training. A half-hour course on accessing the database and entering the appropriate information was sufficient. Older scientific staff required more persuasion. Most of the Agency staff, including library staff, had never touched a terminal. Most senior scientists requested that their secretaries become proficient in entering references into the Order database: a workable solution, since most of the secretaries were already familiar with the VMS wordprocessing system. Secretaries may not have understood why they performed certain commands in a specific way, but understanding was not a requirement as long as they followed prescribed procedures and typed accurately. 'Progress, therefore, is not an accident, but a necessity.' If scientists wanted reprints, they or someone in their department had to learn to enter the request into the automated system. Initially, library staff were inundated with calls for assistance (instructions had been misplaced or there was not enough time to read them), especially when there were problems with the VAX which affected all systems running on it. Library staff eventually became knowledgeable not only in the CDS/ISIS program but also VMS commands and operating procedures. IARC staff were being pulled kicking and screaming into the 20th century.

Success breeds success. Staff who used the automated system raved about it to their colleagues. It was simple to use and ensured follow-up. Benefits to the end-user were straightforward and readily recognisable, in terms of decreased time and effort in tracking references and decreased frustration. Scientists (either directly or through their secretaries) gained untold hours per week which they more profitably spent in research. Of course, the one drawback to the user was the new ability of the library to charge back the cost of the reprint. Library staff

were delighted with the increased ease of managing a centralised ordering system. In spite of increasing orders, the fulfilment function itself was facilitated. Routine sorting/printing routines produced ASCII files uploaded to fulfilment sources such as the British Lending Library, or printed for faxing to local libraries who did not have e-mail. Library staff now had an organised procedure allowing automatic generation of not only listings but management reports. Library staff were able to process three or four times the number of reprints with the same number of staff. Productivity increased, efficiency increased, the budget stabilised and, beyond doubt, credibility in library services soared.

A concomitant benefit was the implementation of a common database for scientists' collection of reprints using the same CDS/ISIS software. Common fields were retained with the Order database where possible. Input was initially performed by library staff. As the demand increased, some scientists provided their own personnel for data entry. Instead of throwing their reprint onto an ever-growing pile(s) on the floor, reprints were now systematically numbered and stored on shelves. Most scientists who participated in the Bibliographic Reprint System (BIB) learned how to access it themselves or requested their secretaries to learn. Some scientists requested that cards be produced, alphabetised by author, continually updated and accessible. The system became especially valuable when urgent requests were required to meet tight manuscript submission schedules (did I already request this article? did I receive it? where did I put it?).

Thus, the Order and Bib databases for interlibrary loans and staff collections of reprints had an overall positive effect on patron behaviour. The speed of retrieving reprints definitely increased, and costs were more efficiently controlled.

2.2. Book orders

The next library function to be automated was the ordering of books. The library purchased 171 books in 1990. This number increased to 274 in 1991. Even though the numbers were small, manual files maintained by title and source of fulfilment were no longer effective or manageable. The paperwork required to process each book was extensive. All books were ordered directly from publishers; requests for status reports or outstanding orders were provided after painstaking research; follow-up was nil. Scientific staff were convinced that the manual procedures were slow, expensive and inefficient. In addition, Author, Title and Subject catalogue cards had to be typed and filed when the book was received: a tedious and time-consuming project. A book ordering system was created using the CDS/ISIS database management system.

The book ordering system was more successful with library staff than with library users, for a number of reasons. Even though the interlibrary loan and book ordering databases were similar, the book database was more complex. Secretaries were more familiar with the former, having typed more interlibrary loans than book orders. Even though detailed instructions were provided, they were not consulted. The software, while easy to use for data entry, required knowledge of not only Boolean operators but command language for searching and printing. Users with no database experience hesitated to learn something that seemed complex and foreign. The amount of instruction did not make any difference. Those staff who could, did, while others called on the Library for help in data entry or to search/print. Library staff, however, found the system not only exciting but more efficient and effective than manual files. Duplication of orders, lost orders or unfulfilled orders were easily identified. Information about each order was now available in detail, describing sources of fulfilment, cost and so on. Costs were easily tabulated for management reports. Library staff processed increasing numbers of orders with less time.

Cataloguing information (e.g. ISBN, Call Number) was entered into the database in addition to ordering information. Catalogue cards by Author, Title and Subject were now produced and printed by sorting/printing routines. Over time it became increasingly obvious that library staff were duplicating the effort: the information already existed in the Book database. Manual catalogue cards were retained mainly for the benefit of staff members who refused to learn how to use automated systems. Library staff decided to discontinue this process and use the online Book system to enhance the manual card catalogue. Library staff were consulted frequently for recently-received books on particular subjects. Even though full automation of the entire catalogue was advertised by word of mouth or by means of the quarterly Library Bulletin, library users insisted on using the outdated catalogue cards.

Scientists suffered during this interim period by not having access to books currently on shelves. Library staff spent considerable time in answering simple reference questions. Even the younger trainees/visiting scientists who consulted books more frequently found using a database management system more difficult than a library online public access catalogue. Users found the displaying and printing of search results cumbersome, especially when knowledge of database commands was required. Whereas reprints formed an integral part of the research effort, books tended to be more academic. Perhaps the very nature of a research institution dictated the success or failure of a specific activity. Reprints formed the lifeline of the research community. In order to publish, one needed access to current information. Or perhaps the real roadblock was the traditional manner of management, i.e. delegating the rudimentary tasks to support personnel. The different levels of expertise were clearly delineated. It wasn't until Bill Gates sat down at his own PC and started typing that the traditional manner of management changed, and it changed faster in the US than in Europe.

2.3. Online public access catalog, serials

Library staff evaluated CDS/ISIS as a means of replacing the Kardex. However, the software could not perform all the required functions associated with a full OPAC. Journal check-in — a three-dimensional model consisting

of receipt dates, publication dates, volume and issue information based on a calendar year — could not be programmed, nor could claim fulfilment. Bar codes were not available. The library thus arranged to lease the Soutron Library System, a VAX program accessible to all personnel through the same terminals, as the interlibrary loan and book ordering systems.

Journals were automated first. Library users were encouraged to use the system to identify journal receipt dates, answer volume/issue queries, and so forth. The library arranged brief training sessions in how to search and print journal information. However, the software was considerably different from that used to order interlibrary loans and books. Library users were resistant to learning yet another set of commands. It was easier to come to the library than query the online system. The OPAC was mainly used by library staff who found the jobs of claiming missing issues, printing reports and managing subscription information much simpler than performing the tasks manually.

Archival library books were entered onto the Soutron Library System after the conversion of journals. A number of library users bemoaned the loss of the traditional musty smell which wafted up when flicking through the cards. Once again, the library was lauded for its initiative and increased efficiency while the usage of library services declined. Once again, temporary staff — i.e. visiting scientists, trainees — were much more enthusiastic about using automated library services than permanent staff. Older scientists were not used to VAX dumb terminals and preferred to delegate this function to support staff.

And then came the turning point. It came at a time when some scientists had already obtained their own Macintoshes or PCs for specific projects. However, use of these machines was haphazard and limited by the scientist's ability to find applicable software independently. Installation often proved a nightmare. The administration changed and this changed the way people worked. PCs were purchased for all secretaries and MS Office, which includes Word, Excel and PowerPoint, was installed. The biggest global change was the standardisation of dissemination of information about systems and software which might have applicability not only to office-oriented tasks but to their installation and maintenance.

3. Continuing automation of the library

In 1988 the IARC Library ran 102 monthly current awareness updates and 725 retrospective searches on Dialog. An analysis of the cost/benefit of searching Medline/Cancerlit in-house on CD-ROMs showed a potential savings of \$8000/year. Two CD-ROM towers were purchased and placed in the library at approximately the same time as initiation of the interlibrary loan and BIB systems. This in-house ability to search Medline not only provided a cost/benefit, but it provided a quicker turn-around in search results and it eliminated problems experienced with modem connections.

Once the in-house availability of Medline/Cancerlit became known, a few scientists expressed interest in using it during evenings and weekends. The library PC was made available on a strict access basis. However, the scientists did not have the time to learn the system thoroughly (e.g. searching MeSH terms) and library staff were not available to give immediate instruction. Searches were performed in a somewhat haphazard manner. It quickly became evident that due to the broad nature of the searches, more information was retrieved than absolutely required. This ineffective retrieval meant more work in weeding out false drops and quality control. Library staff, appalled, assumed that some justification was warranted, considering the increased independence, flexibility and self-confidence in using library services. The number of scientists using Medline was small; however, it was encouraging.

Use of the library PC by end-users resulted in problems: for example the PC was used for services other than strictly Medline searching. At times, even the AUTOEXEC.BAT and CONFIG.SYS files were changed. The library decided to purchase two InfoServer CD-ROM towers which allowed the CDs to be networked and accessed directly on the user's PC or Macintosh. This decision was enforced by the increased workload and the overall management decision to delegate certain library functions directly to the scientists.

'Necessity, the mother of invention.' Genetic research required access to gene sequences. The retrieval of records from the sequence databases at the National Center of Biotechnology Information (NCBI) became available, initially through telnet and then through Mosaic or Netscape. This inspired not only interest but enthusiasm for direct participation in the search/retrieval process. Scientists wanted to input their own queries. Equally as important, they wanted to submit their own sequences to the database. The atmosphere became increasingly positive towards automation. Scientists were encouraged to use their PCs and Macintoshes for more than just analysing data. Multiple licenses were purchased, not only for Reference Manager but for Current Contents as well, and made available to each scientist on his or her PC or Macintosh.

To sweeten the learning process, the library encouraged training in 'fun' kinds of activities. E-mail became a great motivator. When scientists discovered that they could communicate with their colleagues instantaneously (or alternatively, were asked by their peers for their e-mail addresses), they accepted more readily to learn how to type their messages. The learning process differed between scientists. Some ranted and raved, frustrated with their own inadequacies in grasping technical concepts. Others developed their own methods which may not have been the most practical but worked. Netscape became a great leveller. Foreign scientists learned how to access local newspapers on the Internet in their native language. Slowly, over a period of six months to a year, all scientists were performing some functions on their PCs or Macintoshes. 'Education has for its object the formation of character.' Surely the majority of the scientists concluded that things would never be the same again.

Value-added services became indispensable: data was reformatted and transferred across library services,

resources were used more efficiently. Sixteen simultaneous users connected to Medline, an all-time high, and the OPAC became a part of the every-day routine, as demonstrated by the increased activity in borrowing.

It became the library's responsibility to install and maintain the software on individual PCs and Macintoshes. No matter that the librarian had never touched a Macintosh in her life. 'Trial and error' became the key terms for success. Had not the library spearheaded automation by establishing the first 'public' online services? It was certainly to the library's benefit to encourage and enhance automation. The role of the librarian dramatically changed to incorporate not only an educational aspect but a more technical one — that of computer specialist.

The increased independence and flexibility of scientific staff markedly affected library staff. Did this mean that, in the long run, library staff would become redundant? Were they deliberately doing themselves out of a job? The fear lasted only a very short while. The redundant clerical functions now performed by automation were replaced by educational and technical responsibilities. Library staff ordered books on PCs and Macintoshes, learned how to 'surf' the net, became conversant in installing software. They were doing more work than ever before! But the work was different — and more satisfying, once the initial fear wore off.

It was not just the end-user affected by automation. The activities and functions of secretarial staff changed as well as those of computer services staff. Whereas the viewpoint was initially very narrow — for example, VAX applications for biostatistical analysis or wordprocessing — the objectives gradually broadened. Computer services staff found themselves reviewing general purpose software — e.g. Eudora for e-mail, Netscape, etc. — and setting up networks to allow communication with the outside world. Things started to snowball and computer staff were increasingly intertwined with individual demands and requests. Initially they resented this additional workload; in the long run, the benefits gained by the Agency as a whole outweighed any discomfort.

4. Conversion to Windows NT

Even though all the systems worked, and the end-users could do practically anything they wanted, system software was disparate: some systems worked on CD-ROM, others as PC applications, and still others as VAX programs (Figure 1). The networking software, Pathworks, only accessed seven CD-ROMs at the same time. Medline was on 14 CD-ROMs. Thus, routines had to be written to disconnect connected CDs and to reconnect to another set of years. Each time the configuration changed the programs had to be re-written and end-users alerted. The one VAX server which serviced all the users created problems. Peak hours of usage resulted in long delays. Some programs hung up the system, thus creating instability. The library was faced with an increasingly computer-literate user community which became increasingly distraught by the different kinds of systems running on different platforms using different commands. If the PC became too charged and DOS was not able to handle the installed programs, end-users had to reboot to suppress certain programs that took up an inordinate amount of space. If the end-user forgot to reboot, the program did not work.

The Library received news that Win/MacSpins would become available on Windows NT. At the same time, the suppliers of the Soutron Library System advised that the VAX programs would only be maintained through an Alpha box, Open VMS, or through Windows. The CDS/ISIS software was being beta-tested under Windows. Everything pointed to Windows NT and the library reconfigured accordingly (Figure 2).

A schedule was proposed to convert the various library databases to Windows NT. Library users were notified through a seminar to expect changes within the next few months. Most PCs had to be upgraded from Windows 3.1 to either Windows 95 or Windows NT client, to allow communication with the Windows NT server. All library users benefited from the conversion, not only in better management of memory and multiple applications but also in more standardised systems.

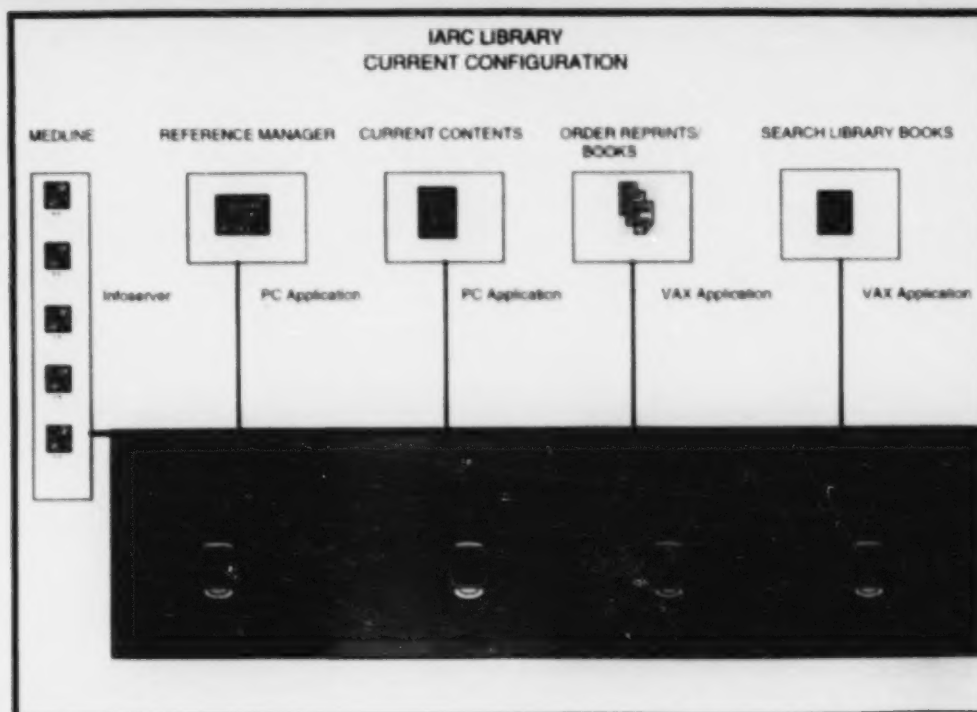


Figure 1: IARC library current configuration.

5. Discussion

Library users initially developed into three types of categories:

- (1) Visiting scientists and trainees formed the most important group of users. They were already familiar with automation and consistently pushed the Agency to advance the capabilities. They embraced the opportunity to grow and expand their base of knowledge. However, they never stayed long enough to impact the permanent staff and new recruits had to be trained repeatedly.
- (2) A second and smaller group consisted of those who did not know automation but accepted it, and tried to adapt to it. They may have used secretarial support in inputting interlibrary loans, but they learned how to search their reprints in the BIB system.
- (3) The third group consisted of older scientists who clung tenaciously to the old ways of doing things. This group formed the majority and relied heavily on secretarial support to bring them up to par with the rest of their colleagues. Eventually, they realised they would have to learn new techniques and technologies if they wanted to keep up with progressive administration and not be left behind. These scientists who wrote their manuscripts longhand clung tenaciously to their VAX dumb terminals.

The learning curve was very slow and painful. Technical difficulties and problems initially caused a great deal of frustration and impeded automation. However, incentives for learning became increasingly available. Direct access to PCs on scientists' desktops facilitated the learning process: scientists could now access Medline, Current Contents and Reference Manager after normal work hours or at weekends. The facility to search Medline and/or Current Contents, export the results into Reference Manager and print bibliographies added a great deal of value to their effort, as did specialised searching, for example genetic sequences.

Macintosh users had different kinds of problems. The programs were one grade behind the PC versions. For example, Reference Manager Version 6 databases had to be converted to Version 5 before they would run on the Macintosh. This meant, at times, the creation of the same database in another version: a time-consuming process, assumed by the library. There was much less expertise in-house to deal with Macintosh problems. WinSpirs users were happily searching for six months before MacSpirs users found the appropriate software which allowed the Macintosh to communicate with the InfoServer towers housing the Medline CD-ROM discs.

Automation increased the prestige of the library in the eyes of library users. The library was seen to be better

organised, more reliable, technically knowledgeable. However, this incongruously did not mean increased use of library services. Some services initially decreased, such as the willingness to live with and access only part of the books as long as the archival card catalogue was available. The library staff were initially more enthusiastic about library automation than end-users.

The conversion to Windows NT was an anticlimax. The hard part was already done. By the time Windows NT came along, all library users had experienced automation to some degree. Life simply became easier with standardised systems, a more stable environment and, hopefully, quicker response times. If the VAX hung up or crashed, library services could still be accessed. End-users would no longer have to use VAX emulators to access the OPAC, interlibrary loan and book databases. No longer would they have to reboot because of memory handling problems or change discs to access various Medline years.

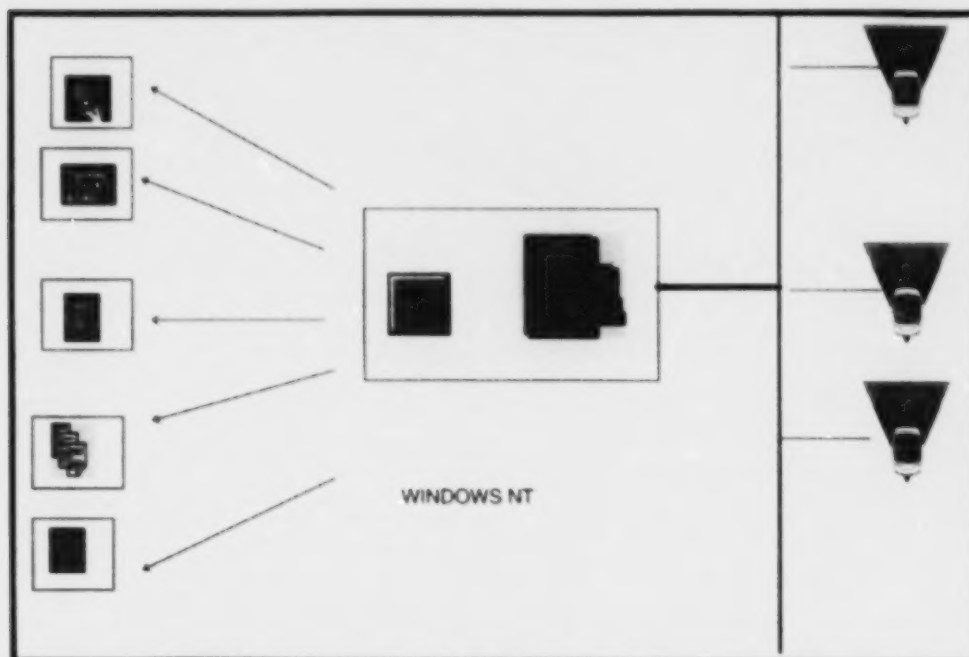


Figure 2: IARC library Windows NT configuration.

6. Conclusions

The degree of success in automating library systems depends on how much the scientist needs the system, the organisational structure and the management style. Initial automation was possible only because IARC operated on a 'laissez-faire' basis. Full-scale automation proceeded as quickly as it did only because administration encouraged, indeed, demanded it.

Training played a smaller part in the success of a system. No matter how many sets of instructions were issued, users preferred systems which worked under Windows where instruction was self-explanatory. Inventiveness played a large part in facilitating the learning process. The end-user recognised the added value of automation.

Adaptation thus depended on previous exposure, need to use the system, ease in using the system, direct incentives, and historical management tenets.

Helis Miido
Librarian
International Agency for Research on Cancer
150 cours Albert Thomas
69372 Lyon Cedex 08
France
Tel: +33 72 73 84 85
Fax: +33 72 73 85 75
E-mail: miido@iarc.fr
http://www.iarc.fr

References

- [1] Faibisoff, S.G. and J. Hurych (1981) Is there a future for the end-user in online bibliographic searching? *Spec. Libr.*, **72**(4), 347 - 355.
- [2] Dixon, R.D. and E.D. Meyers (1982) Initial experiences with an online catalogue at the Boys Town Center Library, *Catalog.Classif.Quart.*, **2**(3/4), 59 - 76.
- [3] Somerville, A.N. (1984) Managing effective information services for end-users in academic sci-tech libraries, *Sci.Tech.Lib.*, **5**(1), 33 - 48.
- [4] Rockman, I.F. and E.E. Pritchard (1985) A comparative analysis of student end-user education and training programmes, *Project*, California Polytechnic State University, Library, San Luis Obispo, CA.
- [5] Rudin, J., N. Hausele, J. Stollak, M.E. Williams and T.H. Hogan (1985) Comparison of In-Search, Scimate and an intelligent terminal emulator in biomedical literature searching, *National Online Meeting Proceedings 1985. New York, April 30 - May 2, 1985*, Learned Information Inc., Medford, NJ, pp. 403 - 408.
- [6] Nicholas, D., Y.W. Pang, G. Erbach and K. Palman (1987) End-users in higher education: the British experience, *Br.J.Acad.Lib.*, **2**(2), 105 - 126.
- [7] Bakken, F. (1994) Global tilgang krever endring av tenkematen. Global access demands change in way of thinking, *DF Rev.*, **17**(6), 157 - 162.
- [8] Dess, H.M., L. Struminger, M.E. Williams and T.H. Hogan. (1996) End-user searching vs. intermediary searching in CAS Online — is the answer in the question? *National Online Meeting Proceedings 1987. New York, 5 - 7 May, 1987*, Learned Information Inc., Medford, NJ, pp. 105 - 110.
- [9] Penhale, S.J. and N. Taylor (1986) Integrating end-user searching into a bibliographic instruction program, *RQ*, **27**(2), 212 - 220.
- [10] Pasanen-Tuomainen, I. (1994) Evaluating the user education, *Inspel*, **28**(4), 473 - 478.
- [11] Barbuto, D.M., E.E. Cevallos (1991) End-user searching — program review and future prospects, *RQ*, **31**(2), 214 - 227.
- [12] Ma, Y. (1994) A comparison of the effectiveness of DAPPOR and a print drill program in DIALOG command language education. A pilot study, *J.Educ.Lib.Inform.Sci.*, **35**(3), 201 - 109.
- [13] Copley, J. (1986) Towards the paperless business library: from print to online business information, *Bus.Inform.Rev.*, **2**(4), 3 - 9.
- [14] Miido, H. (1991) *The Integrated Medical Library*, CRC Press, Boca Raton.



Disintermediation: a disaster or a discipline?

Gillian Allen

Foreign and Commonwealth Office, UK

Abstract: The term 'disintermediation' has developed to describe the finding of information by an end-user without the need for a third party. New developments in techniques for retrieval and dissemination have led to a situation where an end-user can acquire the basic skills necessary for searching the myriad databases available. The information professional needs to become more proactive and add value to the end product in order to meet the challenge of new technology.

The paper concludes with a case study of working practices in the library at the Foreign and Commonwealth Office, London, UK.

Keywords: Disintermediation, database searching, Internet, information provision

1. Introduction

Disintermediation is a term that has been adopted from the economic sector, meaning the investment of money without the mediation of a bank or other controlling factor. In the information profession the term is being increasingly used to indicate that data is being obtained by the end-user without the assistance of a third party.

This is the latest in a chain of events to affect the working practices of the information professional, and might to some perhaps appear as a threat on the horizon. Developments in the past have been enhancements to the professional way of working — the convenience of automated loan systems in the late 1970s and early 1980s; the increased availability of online systems to search ever more remote databases; and the ease with which they can be used, with facilities such as Lotus Notes and the evolution of Windows software in the 1990s. But the presence of these computers and their software is transcending the business environment, and now a vast number of individuals have access to a personal or laptop computer with which they can move their working environment from the office, to the train, to the aeroplane and back to an outpost in a different country — all without the need for an intermediary.

2. A Challenge or a chore?

How should the information professional perceive the new working environment? As a disaster or as a discipline? Enquirers now have a better awareness of the services that they believe they should be offered. The Internet, for example, regularly has column inches devoted to it in the national press, not only in the computer sections or the merger sections in the business pages, but in the general news pages, featuring stories on how it has been used to trap one of the FBI's most wanted criminals in Guatemala, or how a couple on opposite sides of the Atlantic 'met' and fell in love via a chat group. More and more businesses are realising that they can use the Internet for advertising, or electronic ordering. The relatively low cost of an Internet subscription means that more and more people are accessing it at home, becoming familiar with the World Wide Web due to the ease of use of Windows software. One of the aspects which the information professional might consider to be a threat is, how many of these individuals are managing a budget that contains an information professional at work? How many of them are asking 'what does that person do that I can't do?' 'What is value-added about my librarian or information scientist?'

So far, disintermediation sounds like a threat. The threat comes from the availability of computerised systems and the ease with which they can be used. Individuals are able to access networks such as the Internet from their PCs at home — what is to stop them from doing it at their desks in the office? This is where reality has to be faced — a lack of time. Unless one has a dedicated ISDN or Ethernet line, much of the time spent retrieving data from the Internet is waiting for links to be made or data to download. How many people have the spare capacity to wait five minutes for a file to download or print? How many have the time to use a search engine to find new sources of information? The Internet is a vast, continually growing entity — someone has to be able to spend time looking at new Usenet groups or Web sites in case some vital piece of information is missed. It has to be acknowledged, however, that the vast majority of data on the Internet is chaff, but the fact nevertheless remains that somebody has to be able to whittle out the wheat.

Online Information 96 Proceedings

Page 29

3. Training

An important aspect of disintermediation concerns training. Thorough professional training is required to get the most out of an information retrieval system — but individuals can gain a basic knowledge of a system with little or no training. From the financial viewpoint, this is particularly important and may be a reason why CD-ROMs are becoming more significant in the marketplace. Why should an organisation spend money sending an information professional on a specialised training course when other employees can discover the delights of an online or CD-ROM system with simple hands-on trial and error? But will employees want to learn these skills? The University of Westminster spent some time actively encouraging its academic staff to access the Internet, and in their turn, to encourage students to do the same 'but by mid 1995 use had not increased sufficiently to strain the network' (Ref 2).

Gloria Dinerman (Ref 3) has suggested that training takes many forms, from training staff to training the public. Our end-users will need us as intermediaries to instil in them the concepts that will lead to successful and efficient searching. Having said that, what is the difference between a successful search and an unsuccessful one? This is an area that has been closely studied by David Nicholas (Ref 5). His paper studied the online skills of journalists at the *Guardian* newspaper and compared them with those of the librarians. It was found that the journalists (the end-users) were equally as productive as the librarians, although their searches either produced few or many references, while those of the librarians tended to fall in-between. One of his main conclusions was that the end-users required only a basic perception of searching strategies before being able to carry out a search that, in their view, was acceptable.

4. Subject awareness and end-user confidence

The information professional can, therefore, view disintermediation as a potential disaster. I believe it to be a potential discipline, a challenge. Librarians and information scientists will have to become more proactive in their work. They need to seek out new information sources, to recommend them and to deliver timely data to the desk of the person who didn't realise, until that moment, that it was just what he or she needed. Information professionals will need to target the requirements of the workplace more directly and to present the resulting data in such a way that enquirers realise that they would not have had the time to do that investigation themselves.

An important question that has to be raised in connection with evaluation of data by the information professional is, how do individuals know what is of use and what is not? If they already have subject knowledge, this is irrelevant — many information professionals have a first degree in, or prior knowledge of, a subject. Once they become involved in information provision to specialists, this knowledge — if they don't have it already — becomes expected. Evaluation of data, initially in conjunction with the specialist, leads to a development of recognition of gems within the garbage.

This, then is the heart of the matter. This is the challenge which faces the information professional in the final years of the twentieth century. End-users will carry out many of the repetitious searches which the more experienced professional dreads, but when it comes down to needing more detailed data, the end-user will discover that he or she has neither the knowledge or experience required to obtain detailed results. Another term currently appearing in the management press is empowerment. Will disintermediation empower our end-users? I don't believe so. They realise that this simple computer and keypad needs to be in the hands of an experienced individual to get the best from the system, and as a result the information professional gains in status and recognition.

5. The discipline of the future

Earlier I mentioned the fact that the Internet can easily be accessed by individuals at home. But without the intermediaries, how would the end-user find the information required? Searching any database effectively means knowing the best way to do so. Generally there is just one method of doing so. Searching the Internet, however, means using several tools — for example, Gopher, Archie, Veronica, WAIS — all of which use different techniques. What do Lycos, or Yahoo, do if not act as an intermediary? And how did they get onto the Internet? Another individual, or group of individuals, put them there. We are therefore all dependent on each other.

The Internet is the one 'popular' aspect of the information professional's job which has seized the imagination of the person on the street. There are, of course, all the other databases to which we have had access for a number of years — those hosted by Knight-Ridder, FT Profile and MAID, to name just a few. How long will it be before companies currently working with these hosts decide to disintermediate them and mount their data on the Internet instead? I do not see that happening to any great extent, although it is possible that a number of the smaller companies may feel it is easier — and in the long term, more profitable — to mount their own information than pay another organisation to manage it for them. The developments in electronic banking and secure transactions are posing concern to the financial sector, and a consultant from IBM has indicated that disintermediation within the computer suppliers industry will become more widespread over time (Ref 4). In a similar vein, the political and social leverage enjoyed by the media industry through the news and advertising may be on a downward slide as more news becomes available digitally via online delivery systems (Ref 1).

What faces the information industry in the future? Will the software providers use the Internet as their main

platform for development? Intranet systems, mimicking Internet processes, will allow the application of standards like SGML to provide a uniform structure for the provision of information. Neural networks will gradually 'learn' to perform value-added functions which once used to be performed by humans. These developments will become commonplace in companies eager to embrace the technology.

6. Case study

As a case study approach to disintermediation, I would like to outline some of the recent developments in working practices on the enquiry side of the library service at the Foreign and Commonwealth Office (FCO), London.

Since the early 1990s, the FCO library service has developed to embrace a team structure. Four enquiry teams, each consisting of three staff, cover specific geographical regions and functional departments. Three of the four teams also deal with enquiries from British Embassies, High Commissions and Consulates (Posts) overseas. Prior to this team structure, enquiries were dealt with on an *ad hoc* basis. Staff rarely dealt with the same enquirer — or subject — on a regular basis.

The team structure has led to a development of knowledge in subject areas. The librarians dealing with Middle Eastern and African countries, for example, have acquired an awareness of events in those countries and are often able to pre-empt an enquiry by faxing a news article to the appropriate Post or by contacting the relevant department in London. They scan periodicals regularly for articles of interest and are able to produce an SDI service, as well as select books for stock and advise on items to be kept locally at Post. The same developments have occurred for each team. The members of the Management team deal with the information requirements of functional departments and senior FCO management, and therefore have acquired knowledge not necessarily related to specific countries but of personnel issues, arms control, environmental policy and the United Nations, to name just a few subjects.

Staff have also been posted to FCO departments within the FCO in London. There are professionally qualified librarians working in the Diplomatic Service Language Centre, the United Nations documents section and the Research and Analysis Department (RAD). Having started work as the RAD librarian in October 1994 I will outline the work in which I am most involved.

The librarian in RAD is known as the information manager for the department. This means that not only do I act as a liaison point between the main library and the research analysts, but I am also responsible for the departmental computer system and network. The position arose as a result of a Cabinet Office-directed efficiency scrutiny of the structures of both the Library and Records Department and the Research and Analysis Department. The scrutiny recommended the appointment of a librarian to work with the research analysts to cut down any duplication of effort between the output of the two departments.

One of the first tasks I carried out was an information audit. This established the sources of information already used by analysts as well as the sources they were aware of but didn't use, and the reasons for this. As a result, an intensive information awareness strategy was developed, using not only the electronic mail facility already familiar and available on the departmental network, but also a newsletter called *ReAD IT*. This focused not only on training aspects of the network already in place, but also highlighted new information sources available via the Internet and other databases.

The department already had a subscription to an Internet Service Provider (ISP) and had subscribed to a number of listservs. However, a major problem was that the e-mail was checked on an irregular basis, and therefore information was rarely current by the time it got to the end-user. I set up a simple system whereby the e-mail was checked regularly each morning, downloaded to disk and virus-checked before being loaded onto the departmental network and sent to the terminal of the relevant user. The use of the newsletter to highlight database details and sites of interest led to an increased awareness of the availability of the Internet. By mid 1995 there were five standalone PCs around the department being used for Internet access. This did not, however, lead to disintermediation for the RAD librarian. There was an on-going need for training, but once an individual was sufficiently competent to be 'let loose', many found that they did not have to the time to wait for sites to load. This resulted in a greater demand for Internet searches to be done by the librarian. Many sites were regularly updated, thus leading to an increased requirement for proactive consumer awareness.

Regular attendance at Heads of Section meetings were crucial to my own understanding of current hot topics, and meant I could produce items for individual analysts which were relevant to their research topic before being asked. This raised the profile of the librarian's position and in turn increased demand for information from my colleagues in the main library. Analysts were offered organised visits to the library — something that few of them had ever undertaken — and this enhanced their understanding of the levels of assistance that were available.

Visits to external organisations such as BBC Monitoring widened the network of prime sources, and the installation of the Reuters Business Briefing database led to a demand for daily headline listings on specific countries, as well as frequent updates on major news stories of the day. The 'career academics' — many of whom are on secondment to the FCO from universities or other organisations — seem to have a greater need for hard copy to back up their research papers, and will accept and devour information voraciously, while long-established members of the FCO research cadre often use the corporate memory and their exhaustive files to produce equally excellent background research for inclusion in future foreign policy decisions.

7. Conclusion

Librarians and information workers have always been able to adapt their working practices to fit the requirements of their community. We have to accept that others will begin to impinge on what we have always seen as 'our territory', but if we develop our jobs in line with the demands — note I do not say the limitations — of the technology then we should find that disintermediation acts as a discipline and does not become a disaster. Individuals will find that their jobs become far more highly profiled and their contributions to the data handling within their organisations will be marked.

Gillian Allen
Foreign and Commonwealth Office
King Charles Street
London SW1A 2AH
UK
Tel: +44 (0)171 270 0832
E-mail: lib.rad.fco@gt.net.gov.uk

References

- [1] Bell, E. (1996) Pink glow of health fades at FT — the paper's problems are symptoms of a disease facing all nationals, *The Observer*, 21 April.
- [2] Blinko, B. (1996) Academic staff, students and the Internet: the experience at the University of Westminster, *The Electronic Library*, 14 (2), 111 - 116.
- [3] Dinerman, G. (1995) Information professional — a portrait of progress, *Proceedings of the 19th International Online Information Meeting*, London. Learned Information Europe Ltd., Oxford.
- [4] Landry, J. (1996) [Comment from] *Computer Reseller News*, Predicasts Prompt, 17 June.
- [5] Nicholas, D. (1995) Are information professionals really better online searchers than end-users? (And whose story do you believe?) *Proceedings of the 19th International Online Information Meeting*, London. Learned Information Europe Ltd., Oxford.

© Crown Copyright 1996. The views expressed here are those of Gillian Allen alone and not necessarily of the Foreign and Commonwealth Office.



The good, the bad and the ugly: Internet review sites

Alison Cooke

University of Wales, Aberystwyth, UK

Alison McNab

Loughborough University, UK

Betsy Anagnostelis

Royal Free Hospital School of Medicine, UK

Abstract: *In response to the increased use of the Internet for information access and dissemination, and the explosion in the availability of resources which may be of interest to end-users, a number of services are now available which provide access to selected and evaluated resources. Many existing Internet search tools are inadequate due to the high rate of recall, low precision and a lack of filtering. Consequently, the development of evaluative tools is potentially useful for both end-users and information professionals.*

This paper will discuss and compare a number of sites which seek to select, evaluate, review or describe networked information sources. The paper is the result of work carried out by the OMNI Advisory Group for Evaluation Criteria which has been looking at a range of services available via the Internet for accessing selected and evaluated medical information sources, and the criteria they use.

Keywords: Internet, review services, information evaluation

1. Introduction

The Internet initially developed as a global communications channel, allowing the world-wide exchange of messages via e-mail. With the success of the World Wide Web and the widespread adoption of browsers such as Mosaic and Netscape Navigator, there has been an unprecedented explosion in the number of users of networked information as well as in the number of information providers:

'The Internet is frequently described by its users ... as a "realm of pure mind". Anything which is thinkable can be expressed to an audience of hundreds of thousands' (Ref 1).

Commercial organisations now recognise its potential for global advertising and selling. Individual organisations, including universities and other research institutions world-wide, rely on the Internet as a mechanism for disseminating promotional material, course details, research profiles and so on. A number of existing journals, magazines and other publications produce parallel online equivalents and publishers have developed e-zines and online journals where the definitive version is electronic. The media generally has begun to move into the interactive age with online programme information and contact details. In addition the personal home page, listing such materials as individuals' hobbies and interests, pictures of pets and links to pop stars' pages, is a continuing phenomenon of the networked environment. Indeed, the list of available resources is endless.

However, ease of access and dissemination has resulted in a number of problems for network users and recent critics have begun to consider the problems associated with the availability of so much information. Davey Winder suggests that threats to the Internet not only include censorship, pornography and commercialisation but also information overload: 'there's now so much information available on the Net, and so many people trying to access it, that something, somewhere, has got to burst.' Information overload presents two central problems: firstly, 'finding the information you want in a sea of dross', and secondly 'getting connected and managing to download the information once you've located it' (Ref 2). Likewise, Louis B. Rosenfeld identifies similar problems which relate to finding information — determining the existence and location of resources, using the different software available, as well as problems of 'scale' and 'quality':

'Scale: ... the user may win the battle of the Internet, namely, finding a useful information resource, only to lose the battle of traditional information retrieval, namely being overwhelmed by too many hits from a large database.'

'Quality: The majority of the information found on the Internet is lacking in terms of currency, authority and comprehensiveness; in addition, only a handful of resources is refereed in any way. Unfortunately there are few trusted mechanisms for assessing the quality of the Internet's information resources' (Ref 3).

Matthew Ciolek has expressed his concern about the growth of what he calls 'multi-media mediocrity' in an article in which he posits that the ratio of total volume of networked information (measured in megabytes) to information useful to scholars may be much smaller than imagined. He is also concerned with long-term trends and prospects for the quality and reliability of WWW-based information resources (Ref 4). This concern is expressed in part through the creation of the Information Quality WWW Virtual Library, a set of pages which keeps track of online resources relevant for evaluation, development and administration of high quality factual/scholarly networked information systems (Ref 5).

Much research work has been concerned with the development of (WWW-based) search and retrieval tools with the aim of providing effective subject access to resources. Examples include automatically generated indexes such as Lycos and Alta Vista, and manually created 'virtual libraries' or catalogues, such as Yahoo!. Existing search and retrieval tools may assist in the identification of information sources but as the number of resources available via the Internet continues to grow, so the number of resources retrieved becomes increasingly unmanageable. Often a search results in the user being faced with an unwieldy quantity of pointers or links, with no clear way of restricting search results. The user must expend time and energy sifting through results, only to find that material may be out-dated, no longer available, or that a number of individual pointers link to different parts of the same source. Although some search engines offer sophisticated relevance ranking mechanisms and these may improve recall, the first items returned by the search tools are not necessarily the most relevant to the user, and casual users without a background in information retrieval are often not prepared to invest time in learning to use the ranking mechanisms effectively. Furthermore, such practices as word spamming may serve to undermine the possible benefits of relevance ranking.

Prior to the availability of networked information, users could have some confidence in the quality of information available to them. For example, publishers of popular material have a commercial imperative to produce 'quality' material while academic publishers seek to maintain quality through the editing and refereeing processes. However, the Internet enables anyone to become an author and provider of information — authors need not adhere to concerns for an audience but may publish at will. Material may be purely egocentric or of an ephemeral nature and of little interest to others. In particular there is considerably more vanity publishing in the networked environment (for example the personal home page), which has been noted by Hope Tillman (Ref 6). Thus individual users are now faced with the need to identify not only relevant material but also quality material.

In response to such problems, a number of sources and services have been developed which attempt to select, evaluate and review networked information. Journals such as *Online*, *Database* and the *Library Journal* regularly print reviews of Internet resources and *College and Research Libraries News* features an Internet reviews section in each issue. *Ariadne* (Ref 7), a two-year parallel publishing journal project funded under the UK's Electronic Libraries Programme, also includes regular reviews. Such sources may assist end-users but are primarily aimed at the information professional audience.

On the Internet itself a number of reviewing services have been developed. Such services seek to select, evaluate or review networked sources in order both to identify relevant materials and also assess the quality of the material retrieved. The development of such evaluative tools is potentially useful for both end-users and information professionals alike. However, such services are themselves of varying usefulness and quality. This paper seeks to compare a number of such sites, offering a guide to the types and variety of services which are now available.

2. Internet review sites

A number of review services were selected for comparison: A#1 Quality Directory, Cyberstacks, Excite, Healthweb, InfoFilter, Magellan (McKinley), Medical Matrix, OCLC NetFirst, Organising Medical Networked Information, Point Web Reviews, Science and Engineering Network News, Six Senses and WebCrawler Select (see Appendix for URLs). Originally these services were considered with specific, qualitative questions in mind:

- does the site clearly state the *criteria* it uses for the selection and inclusion of resources?
- is the evaluation site selective in the *type of resource* included?
- is the evaluation site selective in the *subject area* it covers?
- if resources are given a rating (e.g. stars or points), are the *criteria for the ratings* stated?
- do resource descriptions/evaluations include the *name of the evaluator*?
- is there any indication of the *expertise/profession of the evaluator*?
- do resource descriptions/evaluations include the *date* on which the evaluation was written?
- is there a mechanism for the *regular updating* of reviews?
- are the *URLs* for resources included in the database *regularly checked for changes*, either automatically by link-checking software or by a human?

- are 'dead' links identified and removed promptly?
- does the site include a search facility?

However, not all sites explicitly provide details about such aspects of their service. Where information is provided it can often be found under the services' FAQ or the 'about' link. For example, both Magellan and Excite include extensive FAQs about their service, staff and criteria for including resources.

In the event, it was possible to draw together a number of common traits into four broad strands: resource selection; the level of description; subject classification and organisation; and the evaluation criteria used.

2.1. Overview of the services

The services selected for comparison vary greatly in terms of the types of resources covered, the subject coverage and the processes of resource selection, description, organisation and evaluation.

Generally the sites may be divided into three groups, noted in Table 1 overleaf.

The first group includes A#1, Excite, Magellan, Point and WebCrawler Select. These are all fairly informal services which have no particular subject focus. No intended audience is explicitly stated or apparent and the guidelines on resource selection and evaluation criteria (where available) tend to be vague and informal. Such services are not produced by those professionally involved in providing access to information, but rather network enthusiasts and 'cyberwriters'. Point, for example, is produced by 'a mix of long time Web fanatics, some professional enthusiasts and even a few students.'

The second group, comprising CyberStacks, InfoFilter, OCLC NetFirst and SENN, is also broad in terms of subject coverage but tends to be more formal. Either information professionals or subject experts are involved in resource selection, description and evaluation; the services are aimed towards the academic or 'serious' information user; and more detailed selection and evaluation guidelines are available.

The last group of four services are those with a subject focus concentrating specifically on health sciences, medical and biomedical sites: HealthWeb, Medical Matrix, Organising Medical Networked Information (OMNI) and Six Senses. With the exception of Six Senses these services take a more formal approach, involving information professionals and subject experts in the provision of services for academic and research users. Formal criteria and guidelines on selection and description are available. Although Six Senses is concerned with a similar subject area it follows a hybrid approach. Even though the review panel includes a registered nurse and a practising clinician, the emphasis of the evaluation criteria tends to coincide more with the services listed in the first group and the rating scheme produces results that can sometimes be difficult to comprehend.

An indicative summary of a range of features offered by the services can be seen in Table 1.

2.2. Resource selection

The review services examined vary in the approach and thoroughness with which new sources are identified and selected. The informal sites, such as Point, are vague in describing the process of resource selection. Point claims that excellence is the only criterion for inclusion and that it 'makes no distinction between commercial, private, or student pages.'

Alternatively, sites such as HealthWeb and OMNI are more specific. These sites primarily involve health sciences librarians and information professionals in the selection and evaluation of sites. Both state that their interest is in providing access to critically-evaluated resources rather than an exhaustive index of health-related information. Medical Matrix seeks to be comprehensive in its coverage of 'global resources that can assist in patient care', but shares a similar concern for quality resources.

The majority of services invite users to recommend sites, although some indicate that these will only be treated as suggestions. For example, Magellan does not guarantee that a suggested resource will be reviewed, only that it will be added to its searchable database. Submissions are subject to an editorial review process and sites are only accepted on the basis of editorial discretion. GNN Select typically adds no more than 1% of recommended sites. None of the sites automatically include reviews submitted by casual browsers.

Table 1: Features of the review services.

	No of sites included	Resource description or evaluation	Rating/points	Subject organisation	Date of review included	Named reviewer
Group 1						
A#1 Quality Directory	unknown	yes	no	yes	no	no
Excite NetReviews	60,000+	yes	no	yes	no	no
Magellan	45,000	yes	yes	yes	no	no
Point Web Reviews	unknown	yes	yes	yes	no	no
WebCrawler Select	2500	yes	no	yes	no	no
Group 2						
Cyberstacks	unknown	yes	no	yes	no	yes
InfoFilter	50	yes	no	no	yes	yes
OCLC NetFirst	50,000	yes	no	yes	no	no
Sci/Eng Network News	unknown	yes	no	no	no	no
Group 3						
HealthWeb	1223	yes	no	yes	no	no
Medical Matrix	2410	yes	no	yes	no	no
OMNI	800+	yes	no	yes	no	no
Six Senses	100+	yes	yes	yes	yes	yes

2.3. Level of description

The level of description offered also varies according to the nature of the service. For example, InfoFilter provides extensive reviews of resources. The reviews are analytical and provide clear guidance on whether a resource is likely to be of interest to a target audience. Other services offer reviews which are brief and merely indicative of the site described while some are simply annotations or descriptions, with little evaluative detail. Only CyberStacks, InfoFilter and Six Senses include the name of the reviewer.

A number of sites employ a numerical rating system. For example, Point rates resources between 0 and 50 for content, presentation and experience while Magellan rates sites by awarding up to 10 points for depth, ease of exploration and 'Net appeal'.

Disappointingly only two services (InfoFilter and Six Senses) explicitly include the date of the review, although an approximate date can be determined in some cases (for example, if the site review is included in a 'what's new' section). Point can be searched in order of review date but this information is not included in the site annotation. In addition to including the review date, the rapidly-changing nature of the Internet requires regular checking of sites to verify links and to ensure reviews are up-to-date. Six Senses claims to re-evaluate sites every two to three months. However, on closer examination some reviews had been written over six months previously and did not appear to have been updated. Medical Matrix states 'entries are continuously updated and verified'.

2.4. Subject classification and organisation

Many services are browsable by subject category. NetFirst offers both Library of Congress Subject Headings and Dewey Decimal Classification numbers. The CyberStacks project has begun to list Internet review services and subject collections by their own classification scheme (Ref 8). Perhaps surprisingly for a review service produced by librarians, InfoFilter has chosen not to include classification codes. NetFirst offers both Library of Congress Subject Headings and Dewey Decimal Classification numbers. Of particular interest to medicine are sites which organise resources by the National Library of Medicine classification codes (NLM) and Medical Subject Headings (MeSH), and here OMNI is a good example.

While most general review sites (such as Excite, Magellan and Point) tend to include a section of 'health/medicine' resources, they do not offer such formalised organisation as the services specialising in biomedical resources.

The majority of review sites are also searchable by keyword. Where a controlled vocabulary is included, such as the use of MeSH, consistency is improved. However, the use of thesauri is not widespread.

2.5. Review criteria used

The nature of the criteria used to evaluate and review information sources is dependent upon the providers of the service. Librarians and other information professionals have long been involved in the critical selection and evaluation of information sources. Evaluation criteria have been developed over the years for the evaluation of reference works, monographs, audiovisual materials and more recently online and CD-ROM sources. Consequently, those services provided by information professionals tend to use fairly formal criteria. For example, CyberStacks bases its evaluations on criteria taken from the American Library Association manual for Reference Collection Development. Resources are evaluated under the broad areas of authority, accuracy, clarity, uniqueness, recency/timeliness, community needs and whether any other favourable reviews are available.

As part of the InfoFilter project, participants are actually concerned with criteria development and a similar set of criteria has been drawn up. Boyd R. Collins, founder of the InfoFilter project states:

'The current reviewing mechanisms we have for print materials arose from the need for objective standards in the face of the mass of materials that are produced yearly. The idea behind the InfoFilter project is to leverage the same well-developed and tested principles of reference book reviewing to the sources available on the Web' (Ref 9).

Likewise, the subject-oriented services use criteria based on those traditionally used by information professionals for evaluating other sources of information.

The 'popular' sites, such as the A#1 Quality Directory, Point and Magellan tend to be concerned with access and presentation issues. Content issues are a consideration (Point considers, 'Just how broad, deep, and amazingly thorough is the information? Are there good links? Good clips? Is it accurate? Complete? Up-to-date?'). However, criteria used in the A#1 Directory include the time taken for a page to load (the service aims to provide links to resources which load quickly using an 'average' computer), pricing, page design and layout and ease of navigation. Likewise, Point asks:

'Is the page beautiful? Colourful? Easy to use? Does it lead visitors through the information nicely? Does it use video, audio and original graphics? Does it break new ground? Is it fun? Is it worth the time? Will we recommend it to friends? All things considered, does this site deliver the goods?'

Magellan:

'Is it well-organised and easy to navigate? Is it innovative? Does it appeal to the eye or the ear? Is it funny? Is it hot, hip, or cool? Is it thought-provoking? Does it offer new technology or a new way of using technology?'

3. User-centred review services: the key to successful evaluation?

The analysis of key attributes of a range of review services highlights a number of issues that bear on the actual utility of such sites. A clear basis for resource selection, usually on behalf of a stated target audience, often underlies the development of services that have applied formalised standards of resource evaluation, description and organisation.

For example, for the casual user the existence of classification codes or subject headings may be of peripheral interest only; but for the searcher who is familiar with the vocabulary structure of their own subject area (e.g. medicine), the availability of a consistent means of accessing this is welcome, such as through the use of a controlled vocabulary.

However it is possibly in the criteria for evaluation that the distinctions among services are most clearly visible.

In most subject-based review services, simple inclusion of resources usually acts as an indicator of quality. In contrast, in the absence of a clear target audience, a number of general review sites (such as Point and Magellan) assign ratings and scores, alluding to objectivity. Yet, more often than not, no special weighting is employed and the figures are simply allocated or uniformly aggregated to provide a magical final score or star rating. This leads to an interesting distinction compared to the subject-oriented evaluative initiatives. Medical Matrix, for example, gives explicit weighting to the evaluation criteria used, which are far greater for issues of scope and content than for presentation and innovation. While the display of ratings and scores and the use of presentation and access criteria may offer the end-user some indication of the nature of a resource, it is questionable how useful such an evaluation can be.

Norman Stevens, writing in 1986, lists 18 criteria that could be considered when evaluating reference books. Yet he also notes:

'Not all of these elements carry equal weight in respect to a single work, and, in some cases, a number of them may not apply at all. Judgement in determining what elements do apply and what weight is to then be given to them is especially critical ... A number of standards used to evaluate a reference book may be of little significance to a user' (Ref 11).

Richard Hopkins further elaborates:

'Any system should be based on the actual information needs of the users involved, not just on available technology or on the subject of the area of study under consideration (expressed in a different way the system must be "user-driven" rather than "technology-driven" or "content-driven") ... Users bring identifiable criteria or priorities to information systems; that is they have certain expectations that they want the information system to fulfil' (Ref 12).

Ease of use, noise reduction, quality, adaptability, time saving and cost saving are examples of these.

In a recent review of issues relating to database information quality, Chris Armstrong reported on surveys of information professionals which were conducted independently by EUSIDIC (the European Association of Information Services) and CIQM (the Centre for Information Quality Management). Both these organisations which invited respondents to rank their views of ten evaluation criteria. The top five rankings were (Ref 12):

EUSIDIC	CIQM
Coverage	Accuracy
Accessibility	Consistency
Timeliness	Timeliness
Consistency	Accessibility
Accuracy	Coverage

The criteria used in the surveys were originally developed during the Southern California Online User Group (SCOUG) 1990 Annual Retreat and are now widely used in the evaluation of online and CD-ROM information sources (Ref 13). The consistently higher ranking of the five criteria above (compared to the remaining five: value, documentation, harmonisation, output and support) might indicate their suitability for weighting any evaluation. While the SCOUG criteria were originally developed and subsequently ranked by information professionals, it is interesting to note how broadly similar they are to end-users' perceptions.

Alison Cooke is currently conducting research into how end-users of the Internet select and evaluate networked information sources. Preliminary results indicate that users are principally concerned with aspects such as the authority, accuracy and currency of information. Information presentation affects some users but the

overriding factor is information content. The limited attention that most commercial services pay to content evaluation would not realistically offer the user much guidance on such aspects as the authority, factual accuracy, reliability or currency of the source, all of which are central issues to consider when evaluating the quality of information.

4. Conclusion

The development of services which seek to select, evaluate or review networked information sources has been a growth area during 1995 and 1996, and a number of sites are now available. Such services differ greatly in terms of coverage, formality and intended audience — sites tend to be aimed either at a general audience, covering a broad range of subject areas and produced informally, or specifically at academic and research users and involving the use of formal evaluation criteria, implemented by information professionals or subject experts. Some services have elected to take a specific subject focus.

The quality of the reviews provided is affected by factors such as the providers of the service and the size of the site/number of resources evaluated. Small sites provided by library and information professionals, such as OMNI, offer detailed evaluations of resources but are more restricted in the number of sites they can cover. Alternatively, the more general and 'popular' services have a broader coverage and are much larger and of interest to the non-specialist, but provide less evaluative detail. In addition the criteria used by such general sites are far removed from the criteria with which information professionals would evaluate other resources.

At present choosing from amongst the evaluation sites seems to be a variant of the precision vs. recall dilemma: such services can either produce a limited number of quality reviews or a large number of reviews resulting from much less rigorous criteria. As discussed previously, the broader, informal sites tend to focus on aspects such as the layout and presentation of sources in order to assess the 'quality' of different information sources, and pay little attention to content. However, such criteria may not only provide limited detail on the quality of sources but may also be of little value to potential users. Thus it is possible to differentiate between the sites in terms of how meaningful the evaluations really are.

The authors of this paper concur with Richard L. Hopkins (cited earlier) that '... the major problems in future information systems will revolve around the processes of reducing the amount and raising the quality of information brought to the attention of the user.' Evaluative sites with little subject focus, or which do not make formal guidelines available, may cover more resources but ultimately users are likely to be concerned with aspects such as information accuracy and reliability, and thus require a detailed and meaningful review of the quality of different sources.

Alison Cooke
Department of Information and Library Studies
University of Wales, Aberystwyth
Llanbadarn Campus
Aberystwyth
Wales SY23 3AS
UK
Tel. +44(0)1970 622146
Fax +44(0)1970 622190
E-mail: aic94@aber.ac.uk

References

- [1] Holderness, M. (1994) Perspective: why did the internet, brain-child of academics and computer specialists become such a significant cultural item? *The Times Higher Education Supplement*, 22 April, 17-18.
- [2] Winder, D. (1995) Threats to the Internet: can the Internet survive? *PC Pro*, July, 162-168.
- [3] Rosenfeld, L.B. (1994) Guides, clearinghouses and value-added repackaging: some thoughts on how librarians can improve the Internet, *Reference Services Review*, 22(4), 11-16.
- [4] Ciolek, T.M. (1996) Today's WWW — tomorrow's MMM? — The specter of multi-media mediocrity? *Computer*, 29(1). <http://www.computer.org/pubs/computer/kiosk/01/kiosk.htm>.
- [5] Ciolek, T.M. (Editor) *The Internet Guide to Construction of Quality Online Resources*, <http://coombs.anu.edu.au/WWWVL-InfoQuality.html>.
- [6] Tillman, H. *Evaluating Quality on the Net*. <http://www.tiac.net/users/hope/findqual.html>.
- [7] *Ariadne*. <http://ukoln.bath.ac.uk/ariadne/>.
- [8] McKiernan, G. *Beyond Bookmarks: Schemes for Organizing the Web*. <http://www.public.iastate.edu/~CYBER-STACKS/CTW.htm>.
- [9] Collins, B.R. (1996) Beyond cruising: reviewing, *Library Journal*, 121(3), 122-124.
- [10] Stevens, N. (1986) Evaluating reference books in theory and practice, *The Reference Librarian*, 15(Fall), 9-19.

- [11] Hopkins, R.L. (1995) Countering information overload: the role of the librarian, *The Reference Librarian*, **49/50**, 305-332.
- [12] Armstrong, C.J. (1995) Database information quality, *Library & Information Briefings*, **62**.
- [13] Basch, R. (1990) Measuring the quality of the data: report on the Fourth Annual SCOUG Retreat, *Database Searcher*, **6(8)**, 18-23.

Appendix: URLs of Internet review services

A#1 Quality Directory	http://www.illuminatus.com/clients/a1source/
Cyberstacks	http://www.public.iastate.edu/~CYBERSTACKS/
Excite	http://www.excite.com/Subject/
Healthweb	http://hsinfo.ghsl.nwu.edu/healthweb/
InfoFilter	http://www.usc.edu/users/help/flick/Infofilter/
Magellan (McKinley)	http://www.mckinley.com/
Medical Matrix	http://www.slackinc.com/matrix/
OCLC NetFirst	http://www.oclc.org/oclc/netfirst/9352nf/netflyer.htm
OMNI	http://www.omni.ac.uk/
Point	http://www.pointcom.com/
Science & Engineering	
Network News	http://www.senn.com/
Six Senses	http://www.sixsenses.com/
WebCrawler Select	http://webcrawler.com/select/



Web search engines: key to locating information for all users or only the cognoscenti?

Nicholas G. Tomaiuolo

Central Connecticut State University, Elihu Burritt Library, USA

Joan G. Packer

Central Connecticut State University, Elihu Burritt Library, USA

Abstract: The proliferation of online services and CD-ROM databases offered by many competing vendors has led to innumerable products with many different interfaces. Information professionals have spent a great deal of time teaching end-users how to use these widely varying interfaces. The World Wide Web is similar to its predecessors inasmuch as it is navigable by search engines provided by various services. The purpose of this research was to determine if end-users, without the benefit of professional instruction, could employ different search engines to locate relevant information on specific topics located on the World Wide Web. Because different search engines use dissimilar search protocols and interfaces, potential problems paralleling those perceived in CD-ROM and online searching were of concern. The results of this study, however, demonstrated that most end-users could effectively use at least two search engines without professional assistance. The reasons for the contrast in end-users' success with World Wide Web search engines compared with problems found in other types of database searching appear to be related to the effectiveness of WWW search engines' software, which can quickly scan a large number of documents and weight the retrieval for relevance. While online database and CD-ROM vendors have made proprietary search syntaxes standard, Web search engines generally provide for natural language searching. The success of users' searches is no longer necessarily contingent on a knowledge of search protocols but rather on the ability to identify correctly the key concepts for which they are searching.

Keywords: end-users, natural language, search engines, searching, students, World Wide Web

1. Introduction

World Wide Web search engines, variously called 'navigation services,' 'Internet directories,' 'searchable subject trees,' 'WWW databases' or 'meta-search engines' are presently the only practical solution for locating specific information in cyberspace. Although numerous print resources offer important information concerning the content of valuable Web sites, these columns, articles, newsletters and directories are static (Ref 1). The World Wide Web's salient characteristic is its dynamic nature. The Web's constant evolution renders even the most authoritative site reviews dated almost as quickly as they are published. Search engines are key to unlocking the Web's potential. As information professionals we correctly view ourselves as a cadre possessing the advanced training and skills required to assist end-users in performing efficient searches of resources of all types (including print, CD-ROM and online). Having participated in continuing education and workshops, digested articles, read monographs and practised our craft for years, we assume that searching is the quintessential blending of art and science that invariably requires ingenuity, experience and thesauri. This is unquestionably true when the online files being searched emanate from, for instance, STN or Dialog. CD-ROMs from SilverPlatter, University Microfilms International and NewsBank, not to mention the dozens of small companies offering compact disc databases (each 'contributing' its own nuances to the searcher/database interaction), require considerable mental agility to ensure efficient searching. Professional searchers scans their memories for the answers to questions such as 'What is the truncation symbol?' 'What is the order of processing of operators?' 'What are the correct field qualifiers?'

Similarly, many Web search engine sites harbour pages of 'Help,' 'Hints,' 'Tips' or 'FAQs' (Frequently Asked Questions) to facilitate relevant retrieval. End-users, especially novices, are often oblivious to these aids. Can end-users expect to locate useful information on the World Wide Web given their untrained approach to searching? This experiment provides a starting point for consideration of this topic.

2. Literature review and comments from Internet discussion list participants

As Nelson stated in 'The lonely life of the end-user,' electronic resources pose significant problems for casual database searchers (Ref 2). Friend referred to 'Interface chaos' (Ref 3). Distress is pervasive among information professionals concerned with end-users' ability to negotiate electronic resources (Refs 4,5,6). Although many articles have discussed training information professionals to use the World Wide Web for research and reference purposes (Refs 7,8,9), few have considered training students (Refs 10,11). Brandt covered teaching end-users to search the World Wide Web from a conceptual perspective (Ref 12). He also authored practical articles addressing instruction in the use of the Web through a text-browser called Lynx (Refs 13,14).

The investigators found no published papers addressing the more specific issue of teaching end-users to interact with Web search engines. A question posted on the discussion lists WEB4LIB (discussion of WWW in libraries) and BI-L (Bibliographic instruction discussion group), in which an investigator requested anecdotal accounts of librarians' perceptions of students' search engine usage, drew a significant number of interesting responses. The following overview of responses reinforces the impression that end-users seldom use search engines effectively. One information specialist commented, 'My consistent experience has been that students are using the search engines the way they use the card catalogue — haphazardly until they find something that works. When they finally ask for help, they are amazed at the hits I am (usually) able to generate and I see them use that information in future searches. What I'm finding with the Internet is that it's taking a lot of one-on-one time to teach students how to search with various engines' (Ref 15). A related remark stated, 'In my observation they [end-users] use Yahoo extensively and the other tools in the most basic and limited way. And do not seem to question exhaustive lists of irrelevant results' (Ref 16). Jessamyn West, Technology Project Coordinator for the City of Seattle (Washington, USA), a role that makes her a primary planner of teaching children and parents how to use computers and the Internet, stated, 'I think there still is the stigma of folks thinking the Internet is for play and therefore a dearth of instruction techniques' (Ref 17). To balance this misconception, she has a Web site document aimed to assist students in using search engines (Ref 18). Another declared, 'This is a special interest area of mine ... With the different interfaces I stress that people should be checking the help screens so that they can take full advantage of the searching capabilities of the search engine. And of course it is important to remind them that the search engines are not mutually exclusive. I recommend using a minimum of three' (Ref 19). Several additional respondents affirmed that they felt making end-users aware of the existence of Web search engines and of their proper usage was essential; they took the opportunity to edify in both formal bibliographic instruction formats as well as at point of use (Ref 20).

Search engines have been acclaimed as the panacea for end-user meandering (Ref 21). Engine providers imply that their services are user-friendly. For example, the top of the Lycos form presently reads, 'Search the Web for' followed by a fill-in box. If accessing Lycos via Netscape Navigator 2.01's NetSearch function, the Lycos form reads 'Find exactly what I want' followed by a fill-in box. The implication is that merely entering words will yield relevant information. InfoSeek's president and chief executive Robin Johnson takes a more explicit perspective. Commenting on a recent advertising campaign promoting his company's services, he said, 'The point of the advertising is that you don't have to be an information retrieval specialist because InfoSeek is like your personal information concierge' (Ref 22).

Information professionals, however, remain concerned that as with CD-ROMs, end-users are not optimally deploying the Web search engines. Is this concern warranted? Herb White commented that for end-users searching is a means to an end, not an end in itself. End-users simply want an answer (Ref 23). The investigators' *a priori* hypothesis was if end-users have historically resisted learning the eccentricities of searching online databases and CD-ROMs, they would probably perform unimpressively using the Web's search engines.

3. Purpose

This experiment attempted to ascertain the degree of success that undergraduates and graduate students, with varying levels of experience using the World Wide Web and Web search engines, and without librarian instruction or intervention, had in locating relevant material on specific topics furnished by the investigators.

4. Methodology

4.1. Selection of participants

The university's electronic bulletin board was used to recruit subjects for the experiment. This ensured that the subjects had some baseline knowledge of networks, electronic mail and computers (i.e. they were reading the notice and replying via computer). Individuals with extensive experience were rejected (e.g. the university's Webmaster had asked to participate and was not allowed). Thirty students requested information on the project. Ten students were recruited and completed the experiment.

4.2. Administration of participant survey and participant orientation

Each participant first completed a brief survey describing her or his previous experience using both traditional electronic library resources, as well as her or his previous experience using World Wide Web search engines. Students were advised of the following: 'A World Wide Web search engine generally refers to a searchable database of files available on the World Wide Web; depending on the coverage and strength of the search engine and the manner in which a search is entered, relevant items may be found at various locations on the World Wide Web. Hundreds of search engines exist; for the purposes of this experiment you will search the following engines: Open Text Index and AltaVista.'

Participants were also told that although the engines basically performed the same function, each may appear different and have its own special features. Investigators stated, 'It is up to you to try to discover some of those features and employ them to do a better search.' No other instruction or elucidation was provided.

Each participant was given seven topics to search. The seven topics were identical for each participant. Investigators mentioned that for each topic valid information existed on the World Wide Web. The investigators began the searching portion of the experiment by having the participant use the 'Open Text Index.' Students were instructed to try one topic at a time. They were further instructed to browse the results and revise their search strategy if they felt it was necessary. Because the investigators were interested in high relevance rather than high recall, the participants were asked to print only the first ten 'hits' from the search they were most satisfied with for each topic. They were reminded not to print the pages of any specific Web site; using an overhead transparency, the investigators illustrated exactly what information should be printed. If zero 'hits' were retrieved, the subjects were instructed to print the page that stated this fact in lieu of a list of sites.

Investigators met with each student individually. Each had a two hour time limit for the appointment. The first half hour was occupied by the orientation and survey. The remaining one and one-half hours could be used for searching.

4.3. Participant searching

Each student was allocated a maximum of 45 minutes to search seven topics on the 'Open Text Index.' Investigators then used a bookmark to switch participants to AltaVista. The same seven topics were searched. Students were again allowed to search for 45 minutes.

Printouts were collected and retained with each student's survey. Data was gathered from the surveys and the search results. Using the printouts, the investigators inspected the students' search strategies as well as their final results.

4.4. Investigator searching

In preparation for the experiment, the investigators attempted to search numerous academic or popular topics sought on the World Wide Web using both engines. Seven topics were chosen. Each topic was adequately covered by both AltaVista and the Open Text Index. Students' retrieval would be compared with 'benchmark' retrieval located by the investigators. The investigators used the available special search features that each engine offered whenever necessary or appropriate to enhance retrieval.

4.5. Success criteria

Investigators determined that a student's search could be deemed successful if they had retrieved the benchmark site(s) that was(were) being used as the reference point(s), or if the retrieval included a site with equivalent or otherwise useful information on the topic, or if the site retrieved contained a hypertext link to one of the benchmark sites.

5. Search topics

The search topics consisted of:

- (1) locate the full text of the Magna Carta;
- (2) locate a listing of Royal Shakespeare Company productions;
- (3) locate Supreme Court Decisions — *but not* Canadian;
- (4) locate information on a press conference given in Madrid by John Major concerning the European Union;
- (5) locate a site containing pictures from the Hubble Space Telescope;
- (6) locate free stock quotes;
- (7) locate information on Poet's Corner in London's Westminster Abbey.

5.1. Benchmark sites and search success scoring

As of the time this paper was written the following sites were established as the 'benchmark sites' (i.e. those sites that had the best information on the topic). Other sites that participants located that had equivalent information were scored as though they were benchmark sites. Sites that mirrored the benchmark sites or had hypertext links to the benchmark sites were scored as benchmark sites. Sites retrieved that had incomplete or only moderately useful material were viewed by the investigators as 'marginal sites' and were scored as being one-half the value of the benchmark sites. If a participant failed to retrieve any sites, found only irrelevant sites or did not complete a search for a topic, a zero was recorded. 'Benchmark sites' were worth two points. 'Marginal sites' were worth one point. All other retrieval or lack thereof was worth zero points.

Full text of the Magna Carta

<http://www.let.rug.nl/~welling/usa/magna.html>

<http://www.traveler.com/~hrweb/legal/magnacrt.html>

<http://www.cs.indiana.edu/statecraft/magna-carta.html> or equivalent sites, etc.

Royal Shakespeare Company productions

<http://www.uktw.co.uk/rsc.html> <http://www.hiway.co.uk/~ei/rsc.html> or equivalent sites, etc.

Supreme Court Decisions — but not Canadian

<http://www.law.cornell.edu/supct/>

<http://www.hplus.harvard.edu/alpha/supcrt.html> or equivalent sites, etc.

John Major's Press Conference on the European Union in Madrid

http://www.fco.gov.uk/current/1995/dec/16/major_pressconference_madrid-summit.txt or equivalent sites, etc.

Hubble Space Telescope pictures

<http://www.physics.carleton.ca/research/Hubble> or equivalent sites, etc.

Free Stock Quotes

<http://www.westergaard.com:8080/goquote.html>

<http://www.li.net/~stanh/stocks.html> or equivalent sites, etc.

Poet's Corner at Westminster Abbey

http://www.demon.co.uk/webguides/london/tourist/to_visit/list.html

<http://trinket.crocker.com/living/berkeley/london/chapter2/westm.html> or equivalent sites, etc.

6. Results

Two points was the maximum score for a successful search. Seven topics were assigned and they were to be searched on two engines. A perfect score equalled twenty-eight points.

6.1. Participants scores compared to self-report of computer use/search experience

All but one participant had used Microsoft Windows. Five participants reported using the World Wide Web 'once in awhile,' four had no experience and one used it 'frequently.' Out of the five that used the Web 'once in awhile' two had found the search engines frustrating while three reported 'sometimes find what I want.' Table 1 shows the scores for all participants as well as their responses to the survey questions. It is interesting to note that the second lowest score was made by a participant who claimed to 'usually find what I want' and stated 'frequently' using the World Wide Web and the search engines.

Table 1: Participants' responses to survey.

Subject	I have used Microsoft Windows	I have used the library's traditional electronic databases (CD-ROM, OPAC, etc.)	I have used the World Wide Web	I have used World Wide Web "search engines" (Yahoo, Infoseek, etc.)	To date, my experience using World Wide Web search engines has usually been:	Search Success Score note: max score possible = 28 (Success % in brackets cf. Table 2)
A grad student	5	1	3	3	3	19 (68%)
B undergrad	2	4	1	1	1	22 (79%)
C grad student	3	2	4	4	4	16 (57%)
D undergrad	5	4	3	3	3	26 (93%)
E undergrad	4	2	3	2	3	22 (79%)
F undergrad	1	3	1	1	1	20 (71%)
G undergrad	4	3	1	1	1	23 (82%)
H undergrad	4	4	3	3	3	20 (71%)
I undergrad	3	2	3	2	2	23 (82%)
J grad student	3	4	1	1	1	14 (50%)

Survey Questions

- I have used Microsoft Windows:
1 = Never, 2 = Rarely, 3 = Once in a while, 4 = Frequently, 5 = Several times each week
- I have used the library's traditional electronic databases (CD-ROMs, OPACs):
1 = Never, 2 = Rarely, 3 = Once in a while, 4 = Frequently, 5 = Several times each week
- I have used the World Wide Web:
1 = Never, 2 = Rarely, 3 = Once in a while, 4 = Frequently, 5 = Several times each week
- I have used World Wide Web "search engines" (Yahoo, Infoseek, etc.):
1 = Never, 2 = Rarely, 3 = Once in a while, 4 = Frequently, 5 = Several times each week
- To date, my experience using World Wide Web search engines has usually been:
1 = I have no experience, 2 = Frustrating, 3 = I sometimes find what I want, 4 = I usually find what I want, 5 = I always find what I want

6.2. Web search engine disparity not a significant factor in results

The differences and similarities of the Open Text Index and AltaVista have been reported (Ref 24). Both provide excellent search capabilities but neither is mutually exclusive with regard to content or search features. Both engines, however, listed the benchmark sites. AltaVista's search protocol is closer to the traditional online or CD-ROM databases. It allows truncation (not automatically but optionally). It allows nested searching. AltaVista's online documentation is comprehensive. But having inspected the printouts, the investigators determined that participants of this study did not capitalise on most of these advantages. Open Text Index is a powerful engine that ignores no stop words. It is extremely straightforward in its presentation.

For either engine, the possible total points that could be scored if all participants had a perfect score on every topic was 140. Participants scored a total of 105 points using AltaVista; a 75% total success rate. They scored 100 total points using Open Text Index; a 71% total success rate.

Ironically, participants who mentioned a preference liked Open Text Index's interface compared to AltaVista's. Although these engines' search screens are dissimilar and have unique features, this was not a significant factor in relevant retrieval.

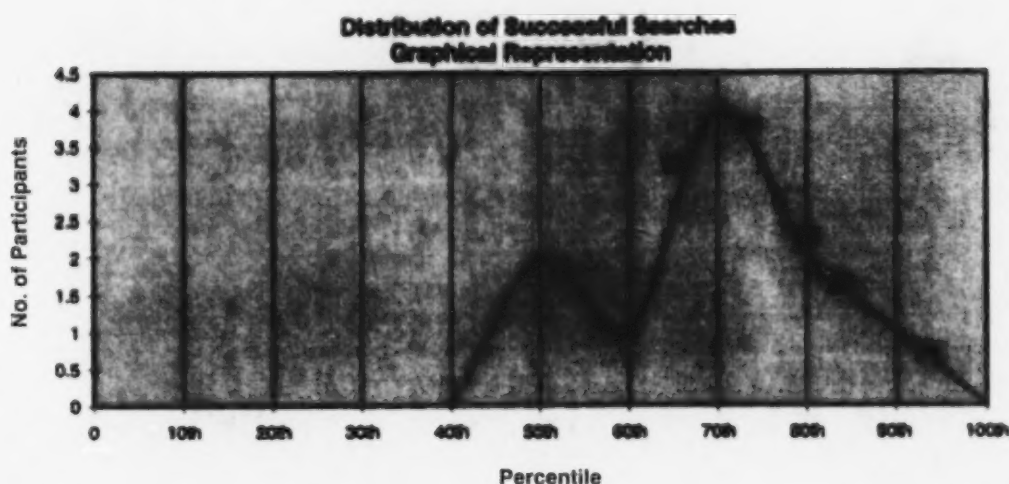
7. Discussion and conclusions

Experience or lack of experience with the World Wide Web or its search engines did not seem to affect the majority of participants' performance. Lack of instruction or unfamiliarity with search protocols generally had a negligible affect on the overall outcome of the end-users' efforts (Table 2). Search engine appearance and disparities in the

search engine's forms had little effect. Notwithstanding the small sample size, the results indicate that end-users of varying experience could use Web search engines reasonably well. The lowest total score was produced by Subject J who accrued fourteen points. Subject J reported having never used the Web and consequently had no experience using search engines. Earning 16 total points, Subject C had the next lowest score. Ironically Subject C reported frequently using the Web, including frequently using its search engines, and 'I usually find what I want.' Aside from these two scores, participants scored acceptably or well, particularly considering they had no instruction and, as evidenced by the printouts, had not used many advanced techniques.

Table 2: Distribution of successful searches.

Percentile	Frequency	Scores
0	0	n/a
10th	0	n/a
20th	0	n/a
30th	0	n/a
40th	0	n/a
50th	2	14,16
60th	1	19
70th	4	20,20,22,22
80th	2	23,23
90th	1	26
100th	0	n/a



On the WEB4LIB discussion list Paul Hollands, Internet Information Officer at Loughborough University, commented that 'The Holy Grail of electronic information retrieval is precision natural language searching.' He qualified this, however, by remarking, 'Unfortunately, having tested Natural Language engines in my previous job which could handle questions like "Give me reports about airbag crash tests that failed," we still found that standard Boolean searching gave much greater precision' (Ref 25). Quint has written that it is natural for end-users to want to search without applying difficult command languages; natural language searching must be realised if end-users are to have the best possible access to information (Ref 26). Within the context of this experiment, natural language searching apparently works as well, in many cases, as the more precise methods that the investigators employed. The scores of the student participants in this experiment bear this out. Students who invoked any of the hints provided by the search engines were in the minority. The search aid most used was AltaVista's asterisk as a truncation symbol. Professional searchers familiar with AltaVista will recall that with each new search a different 'tip' is posted at the top of the AltaVista page. The tips on the use of the asterisk, use of quotation marks for entering phrases and the use of the + (plus) sign to restrict retrieval are frequently rotated. Otherwise students simply picked out the main concepts of the topic and entered them. In several cases, unfortunately, students did not select enough keywords. For example, in some cases students merely entered 'John Major' and clicked the 'Search' button. The investigators had assigned a very specific topic and this strategy was far too broad.

Notwithstanding the overall acceptable outcome of these searches, deficiency in formulating search strategies is a subject warranting comment. In the experimental searches students were provided topics that

were articulated in precise terms. Even this advantage proved insufficient to guide some of their attempts. Omitting keywords (e.g. Westminster, Madrid, European Union) was an apparent problem. Typographical/spelling errors (e.g. Westminster, Abby) hampered retrieval but were never caught or corrected. Consider the student with an imperfect understanding of a topic. In this situation many students may falter in locating useful information. Oberman peripherally supports this: 'A growing number of studies of students and their interactions with electronic databases clearly illustrates the dismal abilities of students, even at the most basic level, of being able to match their subject needs with appropriate computer retrieval systems' (Ref 27). Her comment reflects the idea that the search instruction process must begin at a much more basic level than some information professionals may assume.

Other authors have stated that end-users continue to desire the information specialist's intervention when searching complex topics (Refs 28,29). These authors, however, have not focused their comments on end-users searching the World Wide Web. Many end-users have mounted their own home pages. The Web is available in many homes. Students explore the Web without trepidation. While the Web may be called a database, it is certainly not a database in the same way Chemical Abstracts, Medline, PsycInfo or the electronic Science Citation Index are databases.

We recognise the dynamic nature of the World Wide Web. Its contents constantly change. (Open Text Index altered its 'Simple Search' form midway through this experiment!) P. Steven Thomas, Coordinator of Library Instruction at Illinois State University (Normal, Illinois, USA) stated during the 'teaching students Web search engines' thread on the WEB4LIB discussion list: 'Search engines are evolving as we speak ... Expert systems are going to increasingly take over the online search logic many of us have learned through training and experience. These systems aren't going to replace us (i.e. librarians) but they will greatly simplify the search complexities that have characterised mediated online searching in the past' (Ref 30). As information professionals consider teaching end-users to manipulate Web search engines efficiently, perhaps only the basics require emphasis. End-users should be apprised of the valuable information contained in the hints, help, tips and FAQs that are available for many engines. Information professionals more wisely may choose as their primary goals the satisfactory formulation and refinement of search queries along with the need to choose sufficient, valid keywords. This being done, many end-users may need be led no further than Netscape Navigator's 'NetSearch' button or to a site containing links to Web search engines. Compared with earlier electronic information formats, for example CD-ROMs and 'value-added' online databases offered by various vendors, the efficient nature of World Wide Web search software, with its ability to handle natural language queries, is apparent.

Nicholas Tomaiuolo
Bibliographic Instruction Librarian
Elihu Burritt Library
Central Connecticut State University
New Britain
CT 06050
USA
Tel: +1 (860) 832 2068
Fax: +1 (860) 832 3409
E-mail: tomaiuolo@ccsu.ctstateu.edu

References

- [1] See, for example, Sara Amato's column 'Internet News' in issues of *College and Research Libraries News*. Also see *The Cyberskeptic's Guide to Internet Resources*, Greg Notess's column 'On the Nets' in issues of *Online*, John S. Makulowich's column 'Net Sitings' also in issues of *Online*, Kristin Spence's 'netsurf' column in *Wired* and intermittent articles in *PC Computing*, as well as many library journals.
- [2] Nelson, M. (1987) The lonely life of the end-user, *Wilson Library Bulletin*, **62**(12), 4.
- [3] Friend, L. (1994) Databases by the dozen: the challenge of multiple interfaces at Penn State, *Wilson Library Bulletin*, **69**(9), 38-41.
- [4] Large, A. (1991) The user interface to CD-ROM databases, *Journal of Librarianship and Information Science*, **24**(3), 203-217.
- [5] Rosen, L. (1990) CD-ROM user interfaces: consistency or confusion? *Database*, 101-103.
- [6] Reese, J. (1993) CD-ROM end-user instruction: issues and challenges, *Microcomputers for Information Management*, **10**(2), 131-154.
- [7] Zorn, P., M. Emanoil and L. Marshall (1996) Advanced Web searching: tricks of the trade, *Online*, **20**(3), 15-23.
- [8] Baker, S.L. et al. (1995) Adult services, untangling the Web: the World Wide Web as a reference tool, *RQ*, **34**(4), 441-44.
- [9] Bell, C. (1995) Using the World Wide Web in Libraries, *PNLA Quarterly*, **60**, 29-30.
- [10] Schrock, K. (1995) Surfing made easy, *NEA Today*, **14**(5), 25.
- [11] Ryan, S. and D. Leith (1995) Training with the Web: Internet training in an academic library environment, *Australian Library Journal*, **44**, 22-26.

- [12] Brandt, D.S. (1996) A conceptual understanding of the Internet, *Computers in Libraries*, **16**(3), 39-41.
- [13] Brandt, D.S. (1995) Teaching the World Wide Web with a text-based browser-part one, *Computers in Libraries*, **14**(4), 49-51.
- [14] Brandt, D.S. (1995) Teaching the World Wide Web with a text-based browser-part two, *Computers in Libraries*, **14**(6), 42-44.
- [15] Carroll-Mathes, P. (1996) Teaching students to use search engines [e-mail to N. Tomaiuolo], [Online]. Available e-mail: tomaiulon@ccsu.ctstateu.edu (Note: Ms. Carroll-Mathes's e-mail address is carrolp@sunyulster.edu).
- [16] Moore, P. (1996) Students' use of search engines [e-mail to N. Tomaiuolo] [Online]. Available e-mail: tomaiulon@ccsu.ctstateu.edu (Note: Ms. Moore's e-mail address is moorep5906@aol.com).
- [17] West, J. (1996) Students use of search engines [e-mail to N. Tomaiuolo] [Online]. Available e-mail: tomaiulon@ccsu.ctstateu.edu (Note: Ms West's e-mail address is daspaz@u.washington.edu).
- [18] West, J. (1996) What you give is what you get. <http://weber.u.washington.edu/~daspaz/libr528/libr528.htm>
- [19] Westera, G. (1996) Teaching students how to use search engines [e-mail to N. Tomaiuolo] [Online]. Available e-mail: tomaiulon@ccsu.ctstateu.edu (Note: Gillian Westera's e-mail address is gillian@boris.curtin.edu.au).
- [20] For example, Rise L. Smith, Public Services Librarian at Dakota State University's Mundt Library (Madison, South Dakota, USA) uses a hands-on lab learning environment. In a 50-minute period Smith demonstrates the importance of using 'help,' 'search tips' and 'FAQs.' Students enter a specific search by simply placing terms in the form; but Smith then emphasises the 'power of punctuation.' Students perform the same search on the same engine but are instructed to use the advantages learned from the search tips. Smith generally uses AltaVista or InfoSeek to demonstrate the contrast in retrieval. (Note: Rise Smith's e-mail address is smithr@columbia.dsu.edu)
- [21] Yudkowsky, C. and T. Brock (1996) Internet search engines end aimless wandering, *Washington Business Journal*, **14**(39), 42.
- [22] Elliott, S. (1996) Marketers of two on-line services start playing to the technology-challenged-but-curious, *New York Times*, D9.
- [23] White, H.S. (1995) Information intermediation: a fancy name for reference work, *Library Journal*, **120**(5), 44-45.
- [24] Courtois, M.P. (1996) Cool tools for Web searching: an update, *Online*, **20**(3), 29-36.
- [25] Hollands, P. (1996) Agents (was Re: Students Use of Search Engines) [e-mail to N. Tomaiuolo] [Online]. Available e-mail: tomaiulon@ccsu.ctstateu.edu (Note: Mr. Hollands's e-mail address is p.j.hollands@lboro.ac.uk).
- [26] Quint, B. (1994) Nature's way, *Wilson Library Bulletin*, **68**(7), 61-64.
- [27] Oberman, C. (1996) Library instruction: concepts and pedagogy in the electronic environment, *RQ*, **35**(3), 315-323.
- [28] Fisher, J. and S. Bjorner (1994) Enabling on-line end-user searching: an expanding role for librarians, *Special Libraries*, **85**, 281-291.
- [29] Martin, H. and D. Nicholas (1993) End-user searching at the Guardian, *Online and CDROM Review*, **17**(2), 83-89.
- [30] Thomas, P.S. (1996) Teaching search engines [e-mail to N. Tomaiuolo] [Online]. Available e-mail: tomaiulon@ccsu.ctstateu.edu. In a subsequent communication Mr. Thomas elaborated on his concept of evolving expert search engines, 'My vision is that web users will invoke a search engine that will engage the user in a dialogue of sorts. Such a dialogue would take the form of interrogatives such as ...
 EXPERT SYSTEM: Please input your search question or phrase...
 USER INPUT: What are the advantages of a flat tax?
 EXPERT SYSTEM: Do you wish to limit your search topic to the United States? Y/N
 USER INPUT: Yes
 EXPERT SYSTEM: Would you accept disadvantages as well as advantages? Y/N
 USER INPUT: Yes
 EXPERT SYSTEM: Your search on 'flat tax' and 'United States' and 'advantages or disadvantages' produces 156 results. View them now or save them to a file?
 USER INPUT: View
 ...'
 (Note: Mr. Thomas's e-mail address is stthomas@rs6000.cmp.ilstu.edu)



Java: an explosion on the Internet

Tim Read

Sun Microsystems, UK

Hazel Hall

Queen Margaret College, Edinburgh, UK

Abstract: Summer 1995 saw the release, with considerable media attention, of draft versions of Sun Microsystems' Java language and the HotJava browser. In the past few months Java has been heralded as the latest 'killer' technology in the Internet explosion. Licensing terms have been agreed between Sun and numerous companies including Microsoft, IBM and Netscape.

Java is set to change the way that people interact with intra/Internet. Discussed in this paper are: how a domestic interface development tool became an Internet product; the Java environment and related products: HotJava, Java Virtual Machine, JavaOS and JavaChips; Java's strengths as a developer and user tool: platform independence, programming quality, object-orientation, security, commercial promise and application integration; Java's competitors: ActiveX and Inferno; and Java applications: cybertrade, corporate information flow and information services specific projects. A glossary of terms is provided at the end of the paper. All trademarks and tradenames are acknowledged.

Keywords: application development; information retrieval; Java; network computing; programming languages

1. Introduction: Domestic interface developer to Internet dream product

Java was born as Oak, a programming language to be used for the development of user interfaces for consumer electronic devices such as video recorders, telephones and products of Video on Demand (VOD). Whilst anticipated growth of the VOD market has been slow to materialise, interest in the Internet has surged. Oak became Java and, with little active marketing, software developers rushed to adopt it to write their intra/Internet applications. The attendance record of 6000 delegates per day at JavaOne, the first Java developers' conference held in San Francisco's Moscone Centre at the end of May 1996, demonstrates an unprecedented eagerness for such a young product. For information professionals curious about Java enthusiasm, this paper outlines the programming environment's characteristics and complementary products; its strengths as a software development tool; competitor products and online applications. Inevitably any discussion of this kind necessitates the use of technical terms and acronyms, and for this reason a glossary is appended to the paper.

2. Java environment and products

Since the release of draft versions in Summer 1995 the Java environment has been the subject of an immense amount of media coverage. Several other complementary Sun products which use the Java name have also been developed and subsequently discussed in the computing press. These include the HotJava browser; Java Virtual Machine; JavaOS and JavaChips. To the casual observer the inter-relationships of these products are difficult to distinguish. A discussion of this 'Java jigsaw' serves to put the pieces together in a meaningful fashion.

2.1. Java programming environment

The main product, the Java programming environment, is a programming language (just like BASIC, C or FORTRAN) with a number of additional tools. In its raw form the Java language is unlikely to interest the creator of a library Web page or an average Internet end-user. Computer enthusiasts may take a different view. Tools for creating the small applications in Java known as applets include SunSoft's Java Workshop, Café produced by Symantec, a product of Borland called Latte and Microsoft's Jakarta. Figure 1 shows a short Java program that prints out the phrase 'Online Information 96'.

```

import java.applet.Applet;
public class Small extends Applet{
    public void init() {
        System.out.println("Online Information 96");
    }
}

```

Figure 1: A Java program.

2.2. HotJava

HotJava is a browser. Written in Java, this browser permits on-screen display of the applets. Applets can be loaded from both the user's local disk and, more appropriately, across the intra/Internet. The HotJava application can be considered as a framework into which new units can be plugged dynamically. These units are designed to provide additional functionality for purposes that might not have been considered when HotJava was originally conceived. Users need not be concerned about how the HotJava browser operates, nor worry about compromising the security of their machine.

2.3. Java Virtual Machine

The Java Virtual Machine is software specifically designed to run programs written in Java. Unlike operating systems such as MS-DOS, the Java Virtual Machine creates a 'padded cell' environment for the programs. This prevents virus-like attack of the user's computer.

2.4. JavaOS

The Java Virtual Machine works with the operating system. Although environments such as Microsoft Windows, MacOS or Unix can provide basic functionality for the computer hardware in use, they are not necessarily available for all types of computer such as Personal Digital Assistants (PDAs) like the Apple Newton, and TV set-top boxes for VOD purposes. The JavaOS meets the need to support the Java Virtual Machine by providing very low level hardware routines: threading; memory management; input and output and networking. It can also make provision for higher level class libraries. The overall effect is to provide an identical platform for an application or applet written in Java.

2.5. JavaChips

JavaChips offer a low cost open alternative to the standard processors used in PCs (for example Intel) and workstations (such as SPARC, Alpha and PowerPC). A large proportion of the JavaOS and Java Virtual Machine is implemented by the JavaChip in hardware. This improves the performance of Java applications/applets since the programs do not need to be interpreted by software: they are executed directly. Cheap processor power paves the way for low cost systems such as the network computing devices already previewed by Oracle and Sun.

3. Software developer and user dilemmas — Java's responses

The extent to which a programming environment minimises the software developer's difficulties in the design of applications helps determine the rate of adoption of any new product. Similarly there are criteria of importance to those who want to use the technology to provide services and to the end-user of applications, regardless of language. Java's potential can be assessed against the common problems of platform dependence; programming quality; procedural programming; security and inter-operability.

3.1. Java's platform independence: 'Write once and run anywhere ... safely'¹

The platform for which a computer application is written depends on a number of factors. Most important is consideration of the potential user group: the size of the projected market and competing products within the identified market. Once a decision has been made to write for PC, Macintosh or Unix users, further choices face the developer: the Application Programmatic Interface (API) to be employed; the computer language to use; the object model to adopt. Each of these secondary decisions may limit the market again, or at least necessitate a recompilation or porting exercise to ensure that the application runs across multiple platforms. It is little wonder that less time than the user might hope is devoted to quality assurance and testing.

The biggest benefit of an application written in Java is that it can run on any major platform. The Java Virtual Machine described above provides the same functionality to Java applications regardless of hardware configu-

ration. Windows NT, Windows 95, Windows 3.11, MacOS 7.x, Solaris, HP-UX, AIX, O/S2 and various network devices are all Java-compatible. This represents a huge potential market for the software developer and instead of putting effort into locating a potential client-base, the real priorities of software development emerge: the application's functionality, user interface and quality. The chore of learning and keeping up to date with programming languages for each different environment is eliminated.

3.2. Java promotes quality: reduced chance of program error

Complex and huge volumes of computer code generate modern day computer applications. The smallest error has the potential to cause a machine to crash. C code, on average, contains one error every 55 lines (Ref 1).

Certain features common to other computer languages have been deliberately left out of Java. For example there are no pointers, no goto statements and no need to manage memory explicitly since 'garbage collection' is automatic. By removing these and a number of other unnecessary language constructs, the threat of programming error is greatly reduced.

3.3. Java and object-orientation: new, richer applications

Until the recent adoption of object-oriented technologies programmers have been obliged to 'recreate' or 'rework' earlier code in order to progress a new project. This is analogous to scientists being repeatedly forced to reprove the basic tenet of their subject prior to taking their work any further. With such constraints Einstein might have given up and never produced his General Theory of Relativity!

Java is totally object-oriented so that developers can re-use applets and class libraries that have been developed by themselves and others to create new and richer applications.

3.4. Java security: cynical software, download limits and digital signatures

Downloading programs from the Internet is regarded as a risky business. Users fear viruses, worms and Trojan horses. Ultimately security has to reach the highest level possible without making a system unusable.

With this in mind the Java Virtual Machine, which runs Java applications, has been designed to be sceptical of any code that the user executes. It assumes the worst, checks the integrity of the byte-codes and passes them through a security algorithm. This is to ensure that none of the basic rules set in the specification are bypassed before a program is executed.

Applications downloaded from the intra/Internet are prevented by the Java Virtual Machine from performing many basic functions, such as reading from or writing to the hard disk on the user's machine. Trusted applications or applets installed locally are exempt. Any applet performing in an MS-DOS virus-like manner would normally be obstructed.

Security scares relating to Java have been attributed to implementation errors within the Java Virtual Machine, or very specific circumstances such as situations where cyberthugs have persuaded their gullible prey to save and run unsafe programs locally. Security is under constant review in an attempt to close any loopholes that might be exploited. Digital signatures will give users an additional assurance that download requests are met with the genuine article.

3.5. Java exploits the commercial promise of the network

To a new PC user, landing a personal machine on the desktop is the first priority. Subsequent network connections are a bonus. In the Unix world networking and a client/server approach is a facility that has been taken for granted and exploited by Java technologies.

In many cases Java expects that the classes or components that constitute an applet or application will be loaded from a remote server. The security scheme prevents the application from then communicating with any other system arbitrarily. Remote method invocation (RMI), however, allows interaction with the distant application. One major benefit of this is that little code needs to be downloaded. Another is that service providers are able to hide the actual implementation of what they offer. Examples of commercial uses of networked Java applications include the calculation of insurance policy payments or the execution of online searches based on remote user input. The commercial organisation protects its intellectual property rights to methods in use by limiting downloadable code to algorithms necessary to complete the transaction.

3.6. Java and application integration: 'the network is the computer'²

Computer users rarely ditch older equipment in favour of wholesale migration to a new system. Rather, they cope with piecemeal integration of existing 'legacy' hardware and software. Software developers recognise that user needs are determined to a degree by older configurations. The capacity of software applications to integrate on an equal footing with the *de facto* standard tools of the native environment is key to product success. The volume of developers who adopt the tools as their primary development environment is also important. Currently there are three main approaches to integration.

The current *de facto* standard for integration on the desktop PC is Microsoft's Object Linked Embedding (OLE). This allows users to drop data into applications (OLE drag and drop); link or embed data from one application into another (OLE documents); automate data and function shipping between documents (OLE automation) and supplement the functionality of an application with custom extensions (OLE Custom Controls). The main thrust of this approach is to synthesise business solutions by bolting together 'component' software on

the desktop using a binary application interface standard (Ref 2).

In the Unix environment and beyond the picture is rather different. The network is the focus for application integration. The major players such as IBM, Hewlett-Packard, Digital and Sun conform to the Object Management Group's Common Object Request Broker Architecture (CORBA). Each company implements its own version of this standard. For example Sun uses NEO Workshop, which includes a CORBA compliant Object Request Broker (ORB). Applications components, written in any computer language and to this standard, can then be glued together across the network and upgraded or re-engineered as appropriate.

Straddling the gap between these two approaches are products such as Iona Technologies Ltd's Orbix. Through use of Interface Definition Language (IDL) remote objects are described. The Orbix tools then create an OLE automation server for the PC. This can be called from applications such as Microsoft Excel via Visual Basic. This approach allows considerable freedom to integrate information found from a variety of sources across the intra/Internet.

Developers are striving to break down their software into smaller, more manageable pieces and thereby move away from single monolithic applications. If the interfaces are well defined and conform to standards then users should be able to choose to use and integrate 'best of breed' software packages to fulfil their business requirements. The arrival of the intra/Internet and dynamically extensible products such as Netscape Navigator v 2.0 and the HotJava Browser have broken a potential Microsoft monopoly in this arena.

Microsoft has been moving in the component software direction since the release of OLE v 1.0 but until recently it has been firmly focused on the desktop only. Commercial agreements with Digital, Software AG and others are intended to make OLE/COM available on multiple platforms and, through the use of Distributed Computing Environment/Remote Procedure Calls (DCE/RPCs), network-ready. In September 1993 IBM and Apple announced their equivalent technology called OpenDoc and DSOM which provides similar functionality in a multi-platform environment. This was an attempt to break the Microsoft stranglehold but despite a number of supporters the project results have not met with much visible success or publicity (Ref 3).

When Java was released in mid-1995 both source code and early access products were made freely available on the Internet for both commercial and academic review. It would now appear that Java has leap-frogged OpenDoc as the main competitor to OLE/COM as the preferred option for component software development. For it to succeed on the desktop it needs to provide peer level integration with Microsoft Internet Explorer as the potential vehicle (or 'container' in the industry jargon) of choice. Java Beans fits the requirement of supporting OLE automation, drag and drop and OLE controls. This defines an API for application developers. The completed applets integrate tightly with the environment in which they are executed.

A whole range of new opportunities for the implementation of a corporate desktop results from these developments. It is possible that over the coming years the browser becomes the container for application delivery with information and data being drawn from both the intra and Internet seamlessly into a single interface. Just as the IBM PC supplanted the IBM 3270 terminal as the industry standard desktop, so may the low-cost network computing device slowly erode the dominance of the 'Wintel' PC. This will be welcome news in organisations where computer administration costs are burdensome. A device that is simply plugged into the network is cheaper to service than a suite of PCs (Ref 4).

4. Java's competitors

Developments in this sector of the information industry race ahead at speed. Alliances are formed and products launched, leapfrogged, re-packaged or dropped with alarming haste. The two other major players are Microsoft's ActiveX and Inferno from Lucent Technologies.

4.1. ActiveX

An early Java competitor casualty was a Microsoft Internet Studio product (Blackbird). Now Microsoft's efforts are concentrated on the ActiveX framework. At the same time Microsoft has also committed to embedding the Java Virtual Machine in its Windows products such as Internet Explorer 3.0.

4.2. Inferno

Inferno from Lucent Technologies (part of Bell Laboratories) sets out to meet the gap in the TV set-top box (for VOD), PDA and Internet market. Inferno comprises a programming language called Limbo; a virtual machine named DiS and a communication protocol called Styx. It has already been ported to a number of platforms including Windows NT, Windows 95 and Unix. The success of Inferno may well depend on both its inherent cross-platform functionality and its ability to integrate tightly with native environments.

5. Java applications

Java and the intra/Internet have the potential to change the way people work and obtain information, products and services over the networks.

5.1. Purchase of goods online

One of the biggest opportunities that the Internet offers is the option to sell and purchase services directly online. Those on offer at present are relatively mundane. For example, wine can be bought directly from a supermarket chain in response to a static Internet page advertisement. The platform independence of Java applications means a much enlarged potential market for products. Multimedia animations may be used simply as a gimmick to attract attention to wares for sale, or as an essential component of marketing: for example a video clip can demonstrate product use. Payment may be made using cybercash, an API of the Java Wallet. Nortel has demonstrated how intelligent agents such as software applets can be used in new mobile telephony applications to provide 'all singing, all dancing' user devices (Ref 5).

A route to market that involves much lower overheads than traditional options may be established. This could squeeze the market for established firms and in turn give rise to a new breed of provider of the Yahoo! ilk (Ref 6). With only a Java'tised home page as the shop window the old maxim of 'caveat emptor' may well be in order as there will be little to distinguish the major players from the small time outfits.

5.2. Improved access to corporate information resources

The leading relational database management system (RDBMS) vendors such as Informix, Oracle and Sybase are already racing to provide Java and Internet-ready versions of their tools. In addition to this the development of Java Database Connectivity (JDBC), an API for relational database access similar in style to Microsoft's Open Database Connectivity (ODBC), will facilitate the development and deployment of applications that access corporate databases. Quicker and simpler access to information improves customer service, lowers costs and shortens the period to launch products and services.

5.3. Java applications and information services provision

For the information professional Java has been described as 'yet another tool developed outside the library community that will nonetheless have a huge impact on information services' (Ref 7). Responses to a Net query made in December 1995 demonstrated that American librarians are interested in the benefits that the technology can bring to service provision (Ref 7).

One of the most obvious areas for application of Java in the library context is information retrieval (IR). Applets have already been implemented for some IR purposes. Examples include an applet called Sketch and Fetch that offers an interface between a journal and a chemical database, which was announced in April 1996 (Ref 8); and staff at Cambridge University Press have written code to allow hierarchical searching over subject areas (Ref 9). The Edinburgh Engineering Virtual Library (the EEVL Project based at Heriot-Watt University and funded as part of the British eLib Programme) uses Java as one of the tools for building a gateway to high quality Web resources in engineering. Java was chosen for constructing a customised Web server in preference to C++ because of its ready availability and ease of use. At the time of writing Java applets are being developed to support the searching and browsing options that EEVL provides. They will allow a free floating window containing the options so as to save users the trouble of reverting to the search page within the browser (Ref 10). This approach could be applied to OPAC design to improve interface and search capabilities. Applets incorporating help routines or to set and save user profiles and previous searches improve service to the library catalogue user. National Semi-Conductor built a product catalogue of some 30,000 items with Java that satisfies customers on various platforms quickly (Ref 11).

Java may become an important tool in user education. Multimedia tutorials can aid in developing information skills such as competence in database searching (Ref 12).

It is relatively easy to suggest applications for a new technology in the library environment, and the examples given show that with adequate resources these can be brought to fruition. However, information services personnel face a number of difficulties in integrating technical innovations, and these can limit the adoption and exploitation of new technology which, in ideal circumstances, could serve user information needs. Java might be regarded as a perplexing 'Just-Too-Fast' development (Ref 13) built up by excessive media hype, and library staff in particular may feel cautious in exploring its potential. As Crawford & Gorman (Ref 14) caution '... librarians have been urged to use almost every one of these innovations immediately to protect their libraries from becoming irrelevant' and the result has been library stock rooms littered with equipment defunct after six months of use. Many (often public) cash-poor library services may be obliged to sit and observe developments, unable to afford investment in equipment or training despite enthusiasm for what looks likely to become part of the standard set of tools to access information objects on the network.

6. Conclusions

In the computer industry marketing matters, and any hype certainly helps focus attention on a product. Sometimes, however, it frustrates those who want to keep up to date with developments yet find it difficult to penetrate beyond the hyped messages. At this stage it can be concluded that Java may not be relevant to everybody who follows the computing press or navigates the Internet. However, in the future few will avoid its influence. What Java does offer now is a greater choice of tools for software development for several platforms, easing developer decisions on market penetration and choice of programming language. End-users can benefit

from the enhancements Java brings to Web page design. In the field of library and information science the most obvious applications are in information retrieval. Long term, Java's legacy is likely to lie in the shift towards true network computing.

Tim Read
Principal Systems Engineer
Sun Microsystems
Springfield
Linlithgow EH49 7LR
UK
Tel: +44 (0)1506 672684
E-mail: tim.read@sun.co.uk

Notes

- (1) Dr Alan Baratz, JavaOne Conference, Moscone Centre San Francisco, 29th May 1996.
- (2) Sun Microsystems tag line since 1982.

References

- [1] Klein, R. (1996) Doing the Java jive, *New Electronics*, 27 February, 39-40.
- [2] 'The Microsoft object technology strategy: component software': paper available from <http://www.microsoft.com>.
- [3] 'OpenDoc vs. OLE 2.0 Superior by design — a developer's view': paper available from <http://www.ibm.com>.
- [4] Gartner Group (1993) Do you know what your PC really costs? More than \$40,000, *Industry Service*, 9(12).
- [5] Sugarbroad, I. (1996) 'Using intelligent agents in new mobile applications'. Paper presented at JavaOne, San Francisco, May.
- [6] Kehoe, L. and P. Taylor (1996) The online challenge flowers, *Financial Times*, 14 March.
- [7] Schneider, K.G. (1996) Java: the third wave of the Internet, *American Libraries*, 27(2), 60.
- [8] Rzepa, H. (1996) 'Java Applet: Eqn Viewer and standards in Java applets'. Message to lis-elib@mailbase.ac.uk, 3 April.
- [9] Bursell, M. (1996) Cambridge University Press. Personal communication, 17 June.
- [10] Kirby, R. (1996) EEVL Technical Officer, Heriot Watt University. Personal communication, 13 June.
- [11] Blake, P. (1996) Javascript versus Internet Studio, *Monitor*, 180, 14-16.
- [12] Mulder, C. (1996) Introducing Java to the library world, *MLA News*, March 1996, 11.
- [13] Schneiderman, R.A. (1996) Just-too-fast: why the Internet changes more rapidly than is good for us, *Enode*, 1(1).
- [14] Crawford, W. and M. Gorman (1995) *Future Libraries: Dreams, Madness and Reality*, ALA Editions, Chicago.

Appendix: glossary

ActiveX	Microsoft's competitor to Java
Alpha	Digital chip family
API	Application Programmatic Interface
Applet	Programs created in Java
Blackbird	Microsoft Internet Studio product, now defunct.
Café	Symantec's development environment for writing Java applications
CISC	Complex Instruction Set Computer
Class library	Set of software components for a particular function
COM	Component Object Model
CORBA	Common Object Request Broker Architecture
DCE/RPC	Distributed Computing Environment/Remote Procedure Call
DIS	Virtual machine component of Inferno
DSOM	Distributed System Object Model
HotJava	Internet browser created in Java
IDL	Interface Definition Language
Inferno	Lucent Technology's competitor of Java
Internet Explorer	Virtual machine component of ActiveX
IR	Information Retrieval

Online Information 96 Proceedings

Page 54

Jakarta	Microsoft's development environment for writing Java applications
Java	Sun Microsystems intra/Internet programming environment
Java Beans	A Java API that provides peer level integration with OLE
Java Virtual Machine	A 'padded cell' environment for the running of programs written in Java
Java Wallet	An API for Internet commerce
Java Workshop	SunSoft's development environment for writing Java applications
JavaChips	Pico, micro and ultra Java CPU specifications
JavaOne	The first Java developers' conference held in San Francisco, May 1996
JavaOS	Operating system to support the Java Virtual Machine
JDBC	Java Database Connectivity
Latte	Borland's development environment for writing Java applications
Limbo	Inferno's programming language
NEO workshop	A software development product including Sun's implementation of CORBA
Oak	Java's ancestor, a programming language for domestic appliance interfaces
ODBC	Open Database Connectivity
OLE	Object Link Embedding
OpenDoc	Apple/IBM strategy for component software
ORB	Object Request Broker
Orbix	Iona Technology's MS Windows/CORBA integration product
PDA	Personal Digital Assistant (e.g. the Apple Newton)
PowerPC	IBM chip architecture
RDBMS	Relational Database Management System
RISC	Reduced Instruction Set Computer
RMI	Remote Method Invocation
Solaris	Sun's Unix operating system
SPARC	Scaleable Processor ARChitecture
Styx	Inferno's communication protocol
TV set top box	Control panel for VOD applications
VOD	Video on Demand
Wintel	Windows/Intel



Twenty years of training in online searching: integrating the Internet with the teaching programme

Pieter A. van Brakel

Department of Information Studies, Rand Afrikaans University, South Africa

Abstract: The late seventies saw the disappearance of the dumb terminal, an essential element in the online search process. Consequently, at the beginning of the eighties the popularity of personal computers was putting more intelligence into the online search process. Automatic logon procedures, offline search strategy design, and easier downloading and reformatting of search results were features that shaped the way online searching courses were offered at tertiary level. In the meantime, the high telecommunication costs were replaced by cheaper rates provided by the development of packet-switching and value-added networks. The popularity of online simulations and CAI courses consequently declined because live or hands-on searching by students became much more economically viable. Cheaper communication rates were not the only reason for this phenomenon. Special 'classroom contracts' with very low fees were being made available to tertiary institutions, as well as the creation of small subsets of popular databases, accessible at similar special tariffs. The publishing of databases in CD-ROM format during the second part of the eighties created yet another innovation which influenced the way in which training in online searching was done. Now hands-on practicals could take place on large CD-ROM databases, using a specific host's command language but without being actually connected to the host. The early nineties provided the possibility of networking these CD-ROM databases, thus creating opportunities to run training programmes where multiple students can access the same database to solve the same or different search queries. The Internet and its facilities have brought about yet another need for change. But in contrast with the previous developments, facilities such as e-mail, listservs, FTP and telnet, navigators such as Archie, gophers and the World Wide Web, and Web search engines all contributed to having a profound effect on the whole online scene. This paper will discuss the effect the Internet might have in the way online searching should be taught. Guidelines will be provided to academic staff as to ways and means to integrate these changes into their online search courses: for example the basics of Internet searching, studying and evaluating of World Wide Web search engines, essentials of the HTML language, principles and practices of Web home page design, and rendering information services via the Internet.

Keywords: online searching, Internet, World Wide Web, teaching, home pages.

1. Introduction

Training in online searching techniques inevitably started soon after the advent of large commercial database host systems such as NLM's Medline, Lockheed's Dialog, or SDC. It is common knowledge that when these systems were launched during the mid-seventies, training was not only offered by the hosts themselves but also by the producers of databases. Training at tertiary level followed suit, more or less towards the end of the seventies. In fact, it was at the very First International Online Meeting in December 1977 that Vickery (1977) reported on a BL R&D project meant to develop the use of online services in schools of librarianship and information science in the UK. In those days time was spent on such questions as where to fit online searching techniques into the curriculum (for example, should it be part of the core curriculum?), the exact nature of the instructional objectives, the most effective teaching methods and media, and so forth.

The fact of the matter is that today tertiary training units are still seeking answers to some of the queries raised at that time. The main reason for the attention to the relevancy of such a course was probably that, for the first time, direct hardware and software costs were brought into the teaching process. Training units had, for the first time, to motivate funding for electronic equipment such as terminals, telecommunication, database fees (for hands-on exposure), system manuals, and also for training media developed soon after training started, such as work books, as well as tape/slide presentations and later videotapes portraying the essentials of online searching. Further problems of teaching online searching had to be solved by tertiary training units in the years to come. A decade after the above-mentioned report by Vickery, Large (1987, p. 288) confirmed that although '... online

searching is now an accepted part of the library school curriculum . . . [and] . . . has now entered the mainstream of information studies, reflecting the widespread availability of online systems in the outside world . . . a number of problems continue to plague educators.' His list is representative of the typical challenges that training units have to address:

- *Extend of hands-on experience in searching databases on remote online hosts.* This includes decisions such as the ideal duration of experience, host(s) to subscribe to, types of databases to include and the best database titles to search;
- *Content of a course.* Various topics could be included to strengthen the theoretical part of an online course, for example management and marketing of online services, the economics of the online industry, the role of the pre-search interview and so forth;
- *Which computerised teaching packages to choose?* Several approaches exist to search offline on subsets of databases, for example simulations and emulations. However, even the best examples can never replace online experience on real databases;
- *Cost effectiveness in terms of staff resources.* Because of the costs involved, extensive supervision is necessary when students search live;
- *Location of online searching within the curriculum.* Decisions in this regard proved far from easy: should it be taught as part of a current module, or rather offered as a separate course?
- *Scope of the standard course is steadily growing.* 'It is no longer sufficient to deal with bibliographic databases: full-text and numerical ones must also be covered.'
- *Keeping up to date with developments in a rapid advancing field is impossible.* There is no lack of published sources such as books, journal articles, conference proceedings and newsletters from database vendors which discuss developments such as new search commands, new databases, the latest in electronic equipment and new search approaches.

Surely, apart from faster and more reliable datacommunication facilities, two major technological breakthroughs did contribute to improve the quality of teaching online techniques. The first was the publication of large bibliographical and full-text databases on CD-ROM, and the second the availability of cost-effective local area networks. Now hands-on practicals could take place on large CD-ROM databases published, for example, by H.W. Wilson or KR Dialog, so that practising their command languages online but without being actually connected to the external host systems was possible. The early nineties provided the possibility of networking these CD-ROM databases, thus creating opportunities to run training programmes where multiple students can access the same database during the same practical.

But has the situation changed since the Internet came to the fore? Surely, the various technological and other developments which have taken place since those early days of dumb terminals and slow communication lines put their mark on the nature of online searching, but none of these innovations could surpass the impact the Internet might have on information retrieval techniques, and specifically, ways and means of integrating the Internet (with its facilities and navigators) into the online teaching programme at tertiary level.

A good starting point would be to investigate, as a frame of reference, what online searching constitutes, in an effort to establish which elements of the Internet can be categorised under the term 'online searching'. This will form the basis for the further discussion on how the Internet could be involved in the teaching process.

2. When can 'Internet searching' be regarded as 'online searching'?

An investigation of the publications about the essentials of online searching indicated that not many differences exist between the different discussions on what exactly is meant by the term 'online'. In one of the earlier writings about online searching, Barraclough (1977) explained some of the more important concepts of online:

- *Online processing:* 'Literally it means that the user is in contact with the computer through some direct linkage . . . now [it] is accepted as being some kind of individual terminal attached to the computer either by a direct line or over the telephone network. The next prerequisite . . . is a large . . . store so that the programs for many users can be held within the computer and can be accessed rapidly. The other essential requirement is a means of sharing the resources of the machine among all the users . . .' (Barraclough 1977, pp. 5-6);
- *Interaction:* 'Essentially one is concerned with the reaction of the user to the output produced by the computer and vice versa. It could be thought that it is only the speed at which this occurs that makes the process interactive . . . but another important factor is the amount of skill and intelligence employed by the user during this process' (Barraclough 1977, p. 6);

- *Interface*: '... provide one of the easiest ways of using a computer and indeed ... allow a user knowing very little about computing to be able to access a large computer ...' (Barracough 1977, p. 7);
- *Information retrieval problem*: 'The first problem in setting up a retrieval system is to produce a good definition of the documents to be entered into the database. Clearly all the bibliographical information has to be held. Most systems attempt to reduce the size of the text by either using abstracts or key words ...' (Barracough 1977, p. 7-8).

More recently, in the introduction to a volume on training for online, as part of Taylor Graham's series *Foundations of Information Science*, Jackson (1987, p. 1) provided the following, more recent, description — the emphasis is however still on bibliographic databases, as was referred to in Barracough's 'retrieval problem':

In online searching for the retrieval of information, the searcher at a terminal is connected to the computer which holds a collection(s) of data in machine-readable form known as a database(s). The searcher interacts with the computer and poses a series of queries, in the form of search terms or indexing terms which will correspond with those in the collection of data in the computer store. The computer responds by listing items identified by these search terms. The searcher can then adapt the search according to the response from the computer.

Apart from more sophisticated machine processing, more versatile search software which reflects more powerful search commands, and friendlier user interfaces (for example Windows and Windows-based software such as DialogLink), the modern meaning of 'online searching' did not change one bit from those earlier writings. Full-text databases have been added to the list of sources being made available online, and these and other sources being published in CD-ROM format were distinguished by the term 'local online'. But even 'local online' still comprises the same idea, that is, interactive manipulation of data, bibliographic references and/or the full-text of information sources via a terminal and by an end-user.

The discussion so far can be classified as addressing the 'traditional' view of online searching. But how does this differ from information retrieval activities when using the Internet? The first point to remember is that the Internet provides the facilities for the development of a complete information system (or systems) in its own right. One cannot compare the Internet's navigation facilities with traditional online activities, as described above: that is, manipulating the content of large bibliographic and full-text databases interactively. Wittig & Wolfram's (1994, pp. 636-627) contribution provides a good starting point in establishing how the Internet could fit into the online teaching programme. They argue that where the traditional online systems permitted timely retrieval of information from large centrally located bibliographic databases, the latest Internet-based storage and retrieval tools permit access to decentralised resources.

Koenig (1994, p. 365), in an earlier writing, emphasised the trend that the location of information becomes less and less relevant, '... as we move information around more and more readily.' Basili (1995, p. 459) is more specific in reminding us that traditional online database systems have a star-shaped architecture where the local PC links up with the service-supplying computer (the host). The local computer is completely subordinate to the central computer in terms of both query language and type of service. In a single session only one host is accessed and an online search is thus limited to the databases of that host alone. In the case of the Internet the connection is reticular, because it (the Internet) consists of a constellation of interconnected nodes in which in one network session (for example a World Wide Web's) it is possible to execute a search for information which can move from one node (system) to another.

These differences still conform to the 'online' idea: the Internet's navigator searching — for example Gopher, Archie and the Web — can be regarded as 'online searching', because as indicated in the above definitions, processing is executed in an online mode (albeit via TCP/IP and a client/server architecture). There is also interaction via an interface (browsers, that is), and with the main purpose of solving an information retrieval problem. The Internet should be seen as a fast growing online system, consisting of a wide variety of so-called navigators and search services. It is being accepted that the average tertiary training unit is already incorporating the Internet in its courses (see for example Wittig & Wolfram 1994, p. 630), but the argument here is that certain online-based sections of the Internet should be fully integrated into those areas of curricula which were originally designed for the training of online searchers.

3. Online via the Internet: suggestions for course work

This section identifies and discusses those elements which could be integrated with current online courses. From a didactic point of view, these elements represent the results of the so-called situation analysis, which is always the first phase of curriculum development. The identification of these curriculum elements was based upon the author's experience in teaching online searching techniques to information science students. Articles traced on Internet training mainly covered in-service training programmes.

3.1. Positioning of a module for Internet searching

It is strongly recommended that teaching Internet searching techniques should commence only after the student has a good working knowledge of traditional search techniques on commercial online hosts. This well-structured

environment of database producers, hosts and clients (the players in the online industry) represents the platform necessary to acquire a solid understanding of and the skills for manipulating commercial databases cost-effectively. Apart from a few exceptions the Internet's online environment cannot yet be described as a good example of how an online system should function. Lack of security, copyright protection, lack of privacy and consistency among data resources, and lack of standards and reliability of information sources still prevail. However, Notess (1995b, p. 87) recorded recently that the development of databases of Web pages and sophisticated search engines for Internet resources exemplifies the growth of the Internet as a significant player in the online industry.

3.2. Instructional objectives, learning content and practicals

In 1994 Williams (1994, p. 155) wrote that the Internet has provided an atmosphere or environment for greatly improved online services, that is 'the next generation of online'. This module's overall teaching goal should be formulated in such a way that it conveys not only the idea that although this 'next generation' represents completely new online frameworks, sources and potential clients — and accordingly, new online search techniques — the new generation still supports traditional online services. As a starting point, the first (introductory) objective could be formulated and developed into content and methods as follows:

Instructional objective 1

The student can demonstrate why the Internet can be viewed as a new approach to online searching.

Suggested course content and practical experience:

- (1.1) Origins of the Internet (short history of recent developments; how it is managed and funded).
- (1.2) Basic concepts, such as TCP/IP, client/server architecture (explain differences between typical interactive and client/server computer communication).
- (1.3) Additional facilities which can be applied to supplement the online search process:
 - (a) telnetting to database hosts (implications of remote login; cost comparisons; procedures to follow; selection of important addresses).
Practical:
 - (1) Hands-on experience of using telnet to manipulate external databases such as OPACs, as orientation, hence practising the phases of accessing those hosts with telnet addresses. See Keys (1993) for a good overview.
 - (2) Establishing cost-effectiveness of datacommunication of telnet versus packet switching and/or value-added networks; datacommunication speeds should also be compared.
 - (b) How to use e-mail to receive downloaded search results, or ways and means to communicate value-added information products to clients of information services.
Practical:
 - (1) Learning the use of a specific e-mail package, such as Pegasus emphasising the Windows version, by sending and receiving messages and file attachments.
 - (2) Study case studies set up by the tutor to analyse a specific enterprise's information flow patterns; suggesting how e-mail can improve these patterns to improve the decision-making process.
 - (c) Listservs which are available to be kept updated on new online developments (distinguishing listservs from the popular news groups; important sources of listserv addresses; advantages of creating a listserv in an organisation; functions of the listserv coordinator).
Practical:
 - (1) Search for listserv addresses on specific subjects or scientific fields, by using directories from gophers or Web sites. Use, for example, Kovac's online Directory of Scholarly and Professional E-Conferences (1996).
 - (2) Practise the procedures and commands of specific listservs, for example by subscribing and contributing to and cancelling from a listservs.
 - (3) Be involved in creating and coordinating a local listserv to keep clients updated with developments.
 - (d) Current awareness services (SDI) available via telnet (nature of this new approach, namely electronic document delivery) (such as UnCover) and SDI services combined; different systems currently available; criteria for establishing the best system for a specific organisation (for criteria, see for example Mancini (1996)).
Practical:
 - (1) Searching EDD systems for SDI purposes; comparing search results between EDD and those of online hosts (for example KR Dialog's SAVE SDI command).
 - (2) Researching digital EDD systems (such as Ariel) versus the fax-based approach.

- (e) **Managing Internet search activities** (similar as well as alternative steps were identified by Makulowich (1995): define the area and level of expertise; require clients to complete a request form; categorise the Internet in lay terms for casual users; establish a realistic time for retrieving results; adopt a disciplined, systematic approach in your search efforts; understand your operating platform and the major tools available; maintain a file of important servers and addresses; develop your own bookmarks and home pages; learn the shortcuts).

When comparing the Internet's navigators and the amount of reporting being done in the literature, it is obvious that the World Wide Web is fast becoming the most sought after online system. The reasons are self-explanatory: standards and protocols such as HTTP and HTML have been established, friendly browsers have been developed and improved in the commercial market, inclusion of multimedia sources is supported, versatile hypertext-based organisation of information can take place, and so forth. Furthermore, the quality of the Web's online tools (search engines such as robots, worms and spiders) does provide for reasonable retrieval quality. It is hence suggested that one objective (Objective 2) is devoted to accomplishing a working knowledge of the different navigators, and thereafter an additional objective be devoted to an in-depth study of the Web's search engines (Objective 3).

Instructional objective 2

A student can describe the features of the different Internet navigators and use them to search for relevant information

Suggested course content and practical experience:

- (2.1) Working knowledge of the techniques for searching the different navigators, namely gophers, Archie (including Veronica and Jughead), Wide Area Information Server and the World Wide Web (overview of the history of each navigator; unique features; their scope, strong and weak points; application possibilities; hardware and software needed for setting up a local navigator site).
- (2.2) Variables to evaluate/compare the search and retrieval features of the different navigators.

Practical:

This practical for 2.1 and 2.2 is analogous to learning to use the complexities of traditional online host systems such as command language and search procedures. Search queries are being set up and a student is expected to find, download and compare relevant information retrieved from different navigators.

- (2.2) Variables to compare navigators and establishing when to search which one to use for specific type of information

Practical:

After the necessary theory has been dealt with, the student should get the chance to compare the features of the different navigators. Variables could be applied by using specific types of queries, for example date and/or subject coverage, retrieval and precision ratio, user-friendliness, reaction speed, error handling, reliability, etc.

When studying the evolution of the Internet, clearly demarcated levels of development emerge. Notess (1995b, p. 87) explained these developments as follows:

- First, the Internet was seen as a telecommunications network, with hosts such as STN, KR Dialog and OCLC making their databases available via telnet connections;
- with the growth of Internet-based information, for example from the commercial sector, these players in the online industry began to participate by developing gopher servers or home pages (see for example Notess 1995a, pp. 75-78). Although these pages contained information about their databases and services, direct searching could only take place via a suitable Web interface to a host system (interfaces should not be confused with Web browsers such as Netscape and Mosaic).

As we know, the standard interface to online systems has been an interactive, continuous connection, using a powerful command language and search facilities for effective interactive manipulation of data and/or information in these databases. Traditionally, these systems lack the graphical user interface so common in basic microcomputer packages of the day. With the development of the Web, the standard interface could be in form format (Notess 1995b, p. 88). The user can now fill out a Web form and submit a query in this user-friendly way. These developments, representing a new model of online interaction, are reflected in the following two objectives (3 and 4).

Instructional objective 3

A student can demonstrate the various levels of Web search tools for online searching and can find relevant information on different categories of search queries.

Suggested course content and practical experience:

- (3.1) Study of the home pages of different online search services (interpretation of specific links from these pages, such as FAQs, publications such as manuals newsletters, pocket guides, document delivery

options, categories of databases, commands and search features, help desks).

Practical:

Given a list of the URLs of representative database host sites, the student is expected to study and compare their content for orientation, for example the EasyNet 2.0 Web site (<http://www.telebase.com>). A link in its home page can be used to activate a telnet connection to the real EasyNet gateway.

- (3.2) Status of current Web interfaces to host systems (defining a Web interface; advantages when compared to the line mode approach; specific examples such as OCLC's or Medline's Infotrieve, which offers access to Medline using natural language as well as Boolean searching. The list of bibliographic references is ranked by relevancy and also included is a point-and-click ordering system for full-text copies (Medline over the Web 1996, p. 3)).

Practical:

Investigating the existence and features of this type of interface according to a set of guidelines; hands-on experience of using a Web interface in comparison with the standard interface (via commands and search strings). A good example is NlightN (<http://www.nlightn.com>), a recently developed system using a Web interface to link to a composite of hundreds of databases, including sources as diverse as Medline on the one hand and Lycos (a search engine) on the other (Notess 1995b, p. 88).

- (3.3) Studying of and comparing a wide range of currently available search engines such as Lycos, Yahoo, WebCrawler, AliWeb and Harvest (terminology; different categories of search engines; typical search features; level of friendliness and ease of use; criteria for evaluation; relevancy and precision ratio; coverage of home page sites included in a search; search formats, for example forms). A number of comparative articles have appeared since December 1995, for example Courtois et al. (1995), Kimmel (1996). Jian Liu's *Understanding WWW Search Tools* (1996) is probably the best directory to search engines and is regularly updated through <http://www.indiana.edu/~librcsd/search/>.

Practical:

It is essential to activate an extensive hands-on programme to enable students to grasp fully the extent of difficulties involved when using search engines. An appreciation for the extent of these 'robot-generated databases' should be developed, especially problems which occur because of the high retrieval and low precision ratios. Searches with different difficulty levels and different types of search queries should be arranged. The importance of search result evaluation should be demonstrated.

The above three instructional objectives form together a solid basis for understanding the complexities of searching online via the Internet. Depending on the structure, overall goal and student level of this module, further objectives could be formulated to teach students more advanced topics, taking them a step further than the traditional parameters of online searching. Examples are Intranet features and how an information service could utilise these to improve services to their clients; network security (building firewalls to protect databases from intruders); network management; task analysis of the Web master or Web coordinator; outsourcing Internet/WWW maintenance; Java applications; and so forth. Deliberation of these advanced topics falls outside this discussion.

However, now that the Web has become almost a standard to publish in electronic format, it is obvious that the information specialist should be well-trained in using this environment to communicate with and render services to clients. This implies demonstrating a working knowledge of such important features as HTML, home page design and rendering information services via the Web. Instructional objective 4 is an effort to convey the essentials of 'Web admin'.

Instructional objective 4

A student can effectively apply the facilities of a Web site to support information services to clients.

Suggested course content and practical experience

- (4.1.) Basics of HTML, as well as HTML editors and converters (studying the basic HTML tags as well as different editors (HTML conversion software) and converters which are available to assist in the markup process).

Practical:

Hands-on sessions are essential to grasp the technicalities of using HTML effectively. Practicals to convert plain text should be arranged. One or more HTML editors should be available in the training laboratory (see, for example Carl's list of authoring tools (1995)).

- (4.2) Home page design principles (hypertext as a method to organise large full-text databases; principles of planning a hypertext environment (for example story boarding); the home page as departure point for a whole range of information services; criteria for designing good home page files).

Practical:

Time should be allocated to the methods to design hypertext documents, if the theory and

practice of hypertext was not covered in any of the other Information Science courses. Planning approaches such as storyboarding should be experienced. Group work can be strongly recommended when working on specific projects. Case studies is another approach that works well in letting students design a home page and subsequent links for a specific environment. Prototypes are then developed under the guidance of the tutor, for example to render a current awareness service via the Web, create promotional material to market the information service, develop study guides for client guidance, handle an electronic document delivery service, and so forth.

4. Conclusion

The impression one gets from articles on teaching Internet skills is that one should teach this extensive area as a large entity in its own right. This might be a good idea from a didactic point of view, because such an approach keeps the potential learning content together, with the result that a student or professional (during in-service courses) is exposed to the theory and practice in a neatly packed unit. The disadvantage of this approach is that the learner might see the Internet as something separate from the other professional information handling skills. Gone is the holistic view of information and information sources, with the result that the same professional might manage and use the Internet separately from conventional electronic and paper-based information.

This paper's approach is different from the view discussed in the previous paragraph. Instead, it represents an effort to isolate entities relevant to a specific topic — online searching in this case — and to integrate these with established methods. This same approach has been followed when discussing teaching online searching procedures: that is, to select and integrate relevant Internet procedures with the training programme. The instructional objectives, learning content and practicals were provided as examples, and therefore as a frame of reference for further development by a specific training unit.

Professor Pieter A. van Brakel
Department of Information Studies
Rand Afrikaans University
PO Box 524
Auckland Park 2006
Johannesburg
South Africa
Fax: +27 11 489 2797
E-mail: pavb@info.rau.ac.za

References

- Barracough, E.E. (1977) On-line searching in information retrieval, *Journal of Documentation*, **33**(3), 220-238.
- Basili, C. (1995) Subject searching for information: what does it mean in today's Internet environment? *The Electronic Library*, **13**(5), 459-466.
- Carl, J. (1995) HTML helpers, *Internet World*, **6**(9), 83-86.
- Courtois, M.P. et al. (1995) Cool tools for searching the Web: a performance evaluation, *Online*, **19**(6), 16-32.
- Haygarth Jackson, A. (1989) Introduction, in A. Haygarth Jackson (Ed.) *Training and Education for Online*, Taylor Graham, London, pp. 1-2.
- Keays, T. (1993) Searching online database services over the Internet, *Online*, **17**(1), 29-33.
- Kimmel, S. (1996) Robot generated databases on the World Wide Web, *Database*, **19**(1), 40-49.
- Koenig, M.E.D. (1994) Target 2000: some thoughts and predictions, *Online & CDROM Review*, **18**(6), 364-365.
- Kovacs, D.K. (1996) *Directory of Scholarly and Professional E-Conferences*, 10th revision [online]. Available <http://www.n2h2.com/KOVACS>.
- Large, J.A. (1987) Online searching and the library school curriculum, *Aslib Information*, **15**(11/12), 288-290.
- Liu, J. (1996) *Understanding WWW Search Tools* [online]. Available <http://www.indiana.edu/~librcsd/search/>.
- Makulowich, J.S. (1995) Ten tips on managing your Internet searching, *Online*, **19**(4), 32-37.
- Medline over the Web (1996) *IRLA*, **31**(8), 3.
- Mancini, A.D. (1996) Evaluating commercial document suppliers: improving access to current journal literature, *College and Research Libraries*, **57**(2), 123-131.
- Notess, G.R. (1995a) The DIALOG home page, *Database*, **18**(2), 75-78.
- Notess, G.R. (1995b) New models for online interaction, *Online*, **19**(5), 87-90.
- Vickery, B.C. (1977) The experimental use of on-line services in schools of librarianship and information science, in *Proceedings of the 1st International Online Information Meeting*, London, 13-15 December 1977, Learned Information, Oxford, pp. 195-201.

Williams, M.E. (1994) The Internet: implications for the information industry and database providers, *Online & CDROM Review*, **18**(3), 149-156.

Wittig, C. and D. Wolfram (1994) A survey of networking education in North American library schools, *Library Trends*, **42**(4), 626-637.



The future of online education and training

Tony G. Rose

Hewlett-Packard Laboratories, UK

Abstract: *The climate of higher education has changed radically in recent years. There is now an undisguised attitude of competition between universities that sits uneasily alongside the traditional collegiate spirit. Diminishing funding and overt competition have fostered the development of teaching and learning methods that can reach a wider market and that do not involve expensive commodities such as lecturers or classrooms. Moreover, students' expectations are more than ever shaped by a prevailing culture of 'on-demand' resources, unconstrained by time or geography. But are the economic arguments really that clear cut? Is it educationally sound, cost-effective and reliable? Surely we should first identify the pedagogic principles associated with this new medium, and consider how teachers can most effectively provide an interactive, supportive and critical learning environment.*

Keywords: Online education, training, schools, colleges, virtual university

1. Introduction

Recent developments in media and communication technologies are set to revolutionise the provision of education and training. With modern technology it will soon be possible to deliver lectures, assignments, tutorials, simulations, even the contents of the world's best libraries to anyone in possession of a modem and computer. Furthermore, amongst all these concrete benefits comes another that may be less tangible but is equally important: the shift towards a more student-centred approach. Learners are now freed from the barriers of time and space imposed by conventional classroom teaching, and encouraged to interact with the material in ways that could never be supported adequately by traditional teaching methods.

According to many observers, we are now witnessing the advent of the 'virtual university'. Furthermore, as the virtual university expands, it will ultimately lead to the demise of the traditional campus-based university.

But is this extreme scenario really likely to happen, or is it just another case of media hype fuelled by a few reactionary technophiles? The 'Education Superhighway' is certainly the subject of considerable media attention at present. Indeed, the current expansion of the Internet has been described by some as similar to a gold rush (Abbot 1996), leading to periods of frantic activity as companies and colleges register domain names and set up Web sites.

Amongst all this hysteria there are many serious questions that remain unanswered. For example:

- what are the pedagogic principles associated with this new medium?
- is it educationally sound, cost-effective and reliable?
- what skills do learners initially require, and what skills and qualities will the virtual universities help them to develop?
- how can the teachers most effectively provide an interactive, supportive and critical learning environment?

Without adequate attention to these issues we may find that the real future of online education and training becomes one of information overload, random browsing and solitary confinement.

2. The present situation

2.1. The growth of the Internet

In order to place the claims made about the information superhighway in proper perspective it is useful to start by outlining the present situation.

It is estimated that there are presently around 7 million computers connected to the Internet. By comparison, there are around 660 million telephones in use. Clearly the Internet has a long way to go. But with 80% annual growth, compared to 6-7% for the telephone system, the Internet should catch up by the end of the decade. North America still has the lead, with an estimated 17% of the population of US and Canada having Internet

access by January 1996. Interestingly, they spend as much time surfing the Net as watching videos they have recorded (Abbot 1996). Current estimates of the numbers of users on the Internet vary wildly (anything between 30 and 60 million), but it is predicted that there will be as many as 125 million by the end of 1997 (NCET 1995).

Recent research shows that 20% of UK households currently have a computer, and this number is growing rapidly. By the end of 1996 half of these will have a network connection (DTI 1996). It is estimated that by the end of 1995, around 2000 UK schools were connected to the Internet (HMSO 1995). When compared to developments in the US, this figure sounds somewhat modest. But much of this discrepancy can be attributed to the availability of free local telephone calls in the US, and relatively high set-up costs and tariff levels in the UK (HMSO 1995).

2.2. The development of the educational superhighway

The development of the information superhighway has been precipitated by a convergence of three key industries: telecommunications, broadcasting and IT. Fibre optic cable is gradually replacing copper wire and offers the potential to deliver a wide variety of services, including education. An example of such a network is SuperJANET, which is the world's first educational superhighway. SuperJANET includes a 34Mbit/s data network interconnecting 60 sites across the UK, allowing group working, advanced information services, remote consultation and distance learning (HMSO 1995).

Evidently, the first step on the road to building a credible education superhighway has been the development of the Internet, and particularly the World Wide Web. The next step will be the development of higher bandwidth networks. Investment by cable TV companies will provide broadband network connections to homes and schools in urban areas in the near future. But the education superhighway is not solely concerned with the Internet or the forthcoming broadband networks. Other technologies that will remain an integral part include:

- *CD-ROM*: this will continue to be an essential distribution mechanism for many multimedia applications (networked or otherwise);
- *interactive video*: this technology most closely resembles the type of applications that may become available over the education superhighways;
- *satellite communications*: this will help mobile users and those in remote locations. Reception equipment will enable schools and colleges to receive broadcast material, and when combined with transmission equipment, video conferencing becomes possible;
- *microwave and radio technology*: this will facilitate 10Mbps data transfer and improve access in rural areas.

3. So what does the future hold?

The current wave of euphoria (and investment) surrounding the education superhighway would certainly suggest that its arrival is imminent. The logic would appear to be simple: if high quality, properly designed learning materials can be downloaded to anyone in possession of a modem and computer then it would appear that the days of the traditional university (at least, in its present guise) are truly numbered.

3.1. An 'on-demand' service

The simplest and most profound effect will be the elimination of constraints on time and space. Learning materials can be downloaded at any time, to any place. Indeed, the trend toward this type of service is reflected by similar developments in the home entertainment business (such as video-on-demand). Part of this development will be a change in emphasis as learning becomes a more student-centred process. Instead of the learning process revolving around the university, the student will become the focus of the experience (Hutchison 1996).

It is likely also that the student body will expand and diversify, as easy access to the Internet will enable more sectors of the community to engage in learning programmes. These will replace some of the more traditional TV-based programmes, and the capacity for interaction will facilitate a range of learning styles. Furthermore, since the student is in control of the pace of learning, the virtual university may more easily accommodate variations in student ability.

3.2. Changes to the universities

With restrictions on time and space lifted, the virtual university can take greater numbers of students, almost without limit. So what need now is there for so many classrooms and tutors' offices? Perhaps expensive commodities such as these will be the first to be sacrificed to fund the new investment. Universities that are slow to change may find themselves becoming merely nodes on the network, providing laboratory facilities that are accessible via remote links. Alternatively, they could adapt to become centres of research and postgraduate training.

3.3. Choice of modules

The virtual university will inevitably increase student choice, since individuals can now select modules from more than one provider. Why take all your modules from Anytown College when you can download a combination from Oxford, MIT and the Sorbonne? However, such combinations are not totally without constraint: some regulation will still be required in order to be eligible for a recognised qualification.

As universities race each other to provide learning materials for this expanding market, they will inevitably experience a considerable lead time in converting their conventional materials to online programmes. They may then find their interests (and perhaps also their finances) best served by acting as 'learning agencies', presenting materials developed by other organisations alongside their own. Course material from a range of providers can then be channelled into a cohesive programme with a recognised degree title.

But wherever there is selling, there is competition. Will the inevitable market forces lead to excellence, or simply cost-cutting by the providers? Moreover, what financial model is appropriate for such services? Should we be paying for the course, the module, or on a pounds-per-page basis? Since the learning is on-demand, like electricity from a socket, perhaps the payment method should follow suit (i.e. pounds per unit consumed). Other models, e.g. payment by the hour, may encourage learners to cut corners when under financial pressure. Evidently, finding the right approach requires detailed consideration of the political issues as well as the technical and pedagogic.

3.4. Administration

The education superhighway will bring several benefits in the administration of student progress. It becomes possible to monitor students' learning behaviour, logging their activity, recording their choices and interactions. With online submission of course material, hand-in rates can be monitored and warnings issued when individuals are persistently late. Moreover, now that the course material is essentially a single electronic copy, it is much more easily updated than the traditional collection of papers, OHTs and textbooks.

3.5. Automated assessment

Online education enables a variety of automated assessment techniques to become viable. The technology for marking multiple choice tests already exists, and marking schemes can be devised that allow a surprising degree of sophistication. Furthermore, it is even possible to automate the exam-setting process itself. Webmark (Attree 1996) is a CGI Perl application that will create a Perl script to set and mark a multiple choice questionnaire. Students can then remotely sit the test, and the results can be fed back to the student and/or tutor.

But not every academic subject is adequately assessed using multiple choice questionnaires. For some, the essay remains the primary vehicle of expression. However, the day when essays can be marked remotely by machine remains (perhaps thankfully) somewhat distant.

In between the multiple choice question and the complete essay there remains a spectrum of further interactions with potential for automated assessment. For example, students on programming courses typically have to produce examples of their own program code. This too can be subject to online assessment, using both static and dynamic metrics. Indeed, the evidence suggests that students actually prefer to be marked by machine rather than by another person (Benford *et al.* 1994).

But what of other types of student, whose output does not usually consist of program code or multiple choice answers? Is there any way to assess their output automatically? Some work is already underway in this area. Allott *et al.* (1995) have devised an automated marking system that makes a binary decision of truth or falsehood of single sentence exam questions using an activation passing network (APN). Preliminary trials using this system show great promise. However, a major problem with this approach is that the knowledge in the APN has to be hand-coded. Consequently, current research effort is directed toward automating this process by extracting relevant knowledge direct from textual sources (Bowden *et al.* 1996).

Another intriguing possibility is the use of virtual reality as a vehicle for assessment. Whereas written exam papers can demonstrate learners' descriptive understanding of a subject, a VR-based examination could potentially engage their real-time problem-solving ability, involving them as interactive participants in a task that requires specific cognitive skills and abilities.

In addition to the 'technology push' in this direction, there is a political precedent. At the last Labour Party Conference, Tony Blair promised to connect every school and library (permanently) in return for securing BT's future. In addition, every child will have access to a laptop. However, those children for whom the laptop is an everyday tool will not wish to relinquish it easily at the examination door. After all, calculators are now accepted as an essential item of equipment for certain subjects: why not computers?

3.6. Changes to the lecturer's role

In schools, the role of the teachers will broaden to include less authoritarian practices. Consequently, the dividing line between them and the learners will become progressively more blurred as teachers and learners become 'fellow browsers in the cybernetic library' (Pickering 1996).

Similarly, the role of lecturers in higher education is also set to change. The research element may continue as before, perhaps benefiting from improved external links and electronic libraries. But with no classrooms, what new teaching skills will be required? One possible outcome is that lecturers will become courseware designers, developing the multimedia authoring skills necessary to create their own materials. Tutoring can still be done,

albeit remotely. Acting like freelance agents, tutors can be hired on an 'as-needed' basis. Such an arrangement is already commonplace for many Open University tutors (indeed, the OU has pioneered many developments in this area, such as virtual summer schools and 'paperless' courses provided over the Internet). This pattern of working is reflected in society at large, with increasing casualisation of labour and a shift toward contract-based, part-time employment.

4. Some cautionary thoughts

One of the major problems facing the development of a credible education superhighway is the lack of a suitable financial model. It is not clear how best to combine the (possibly conflicting) goals of student choice, quality and return on investment for the provider. Without careful planning, the tendency from both sides (learner and provider) will be to cut corners and costs. A related problem that is particularly acute for students of the virtual university is that since so much material is now held in digital form, the potential for copyright infringement (and its consequent effect on royalty payments) becomes so much greater. The ease with which individuals can cut and paste networked material presents a considerable challenge to a human assessor, and invites a reappraisal of the boundary between plagiarism and legitimate research.

Perhaps the very concept of 'ownership' needs to be reconsidered, now that material can so easily be acquired, processed and re-arranged. On a more pragmatic note, it is possible that our perception of what constitutes an 'original' piece of student work needs to change. Instead of requiring a who's, self-penned exposition, perhaps we should welcome a collection of original and derivative material, on the understanding that its combination represents the student's view of an effective criticism of the subject.

There remains the problem of the 'dark' side of the Internet: viruses and pornography. The latter problem is particularly acute for schools and a number of countermeasures are currently being investigated. One solution is to use an access provider that monitors its own content, e.g. BT's CampusWorld. Full Internet access is allowed but users are encouraged to use the 'walled garden' — a safe subsection of the Internet that focuses on educational material and support. Another approach adopted by some access providers is to exercise a kind of censorship by blocking access to certain Web sites that are deemed 'unsuitable'.

As ever with IT investment, developing an overall strategy remains problematic — how can long-term plans be made when the technology changes so rapidly? Much of the equipment present in UK schools is greater than five years old, and even with expenditure on IT hardware at around £150 million pounds it cannot keep up with the pace of current developments (DTI 1996). But plan they must, since parents have invested massively in educational technology in the past two to three years, in the expectation that their children's learning will benefit. If schools' IT plans do not do reflect this, their annual pleas of institutional poverty will not be received sympathetically (Heppell 1996).

Costs remain a major barrier. A connection to JANET or SuperJANET may be possible for schools sponsored by a higher education institution, at around £100 p.a. for dial-up access and £1000 p.a. for affiliate status. In particular, schools need flexible tariff structures that avoid the open-ended costs usually associated with dial-up charges.

In addition to the hardware costs there are the staff costs. Sophisticated IT skills are required to administer e-mail accounts, newsfeeds, local information servers and network connections. However, in 1993/4 up to 50% of secondary schools and 30% of primary schools had received only initial awareness training or none at all (DTI 1996). Moreover, a significant percentage of schools do not have LANs in their computer labs (Parker 1994).

5. Conclusions

When Gutenberg invented the printing press it helped generate an information explosion that precipitated the Renaissance and in turn led to the establishment of modern science (Foremski 1994). Will the second information revolution have effects of similar magnitude?

There is no doubt that the Internet, even in its present form, has the potential to revolutionise education and training. It already contains much information that is of great value, although much of it is incomplete or hard to locate. However, advances in user interfaces and search tools will go some way to alleviating such problems.

Certainly, the Internet has its faults, notably the lack of formal structure and the absence of guarantees for its contents. But a more charitable appraisal would acknowledge the generosity and altruism of the contributors, and view that as the foundation of a global partnership in learning and cultural development.

There is a still long way to go in building the education superhighway. There are many people throughout the world without telephones, let alone Internet connections. We are thus faced with the possibility of a gap between those who benefit from the Internet and those who do not, both within and across cultures (Pickering 1996). However, the first steps have now been taken and the direction looks irreversible. The infrastructure is developing that will provide high-speed affordable access to schools, colleges and communities. What we need now is the political will to provide the educational services that the next generation deserves.

Dr T.G. Rose
Senior Lecturer
Department of Computing
Nottingham Trent University
Nottingham NG1 4BU
UK
Tel: +44 (0)115 9486017
Fax: +44 (0)115 9486518
E-mail: tgr@doc.ntu.ac.uk
<http://www-uk.hpl.hp.com/>

References

- Abbot, C. (1996) Cerf and the Net, *Times Educational Supplement IT Review*, 5 January 1996.
- Attree, J. (1995) Webmarkhow to create a multiple choice marking script, <http://piscis.sbu.ac.uk:80/BE/SUDP/W3/webmark.html>
- Allott, N., P. Fazackerley and P. Halstead (1994) Automated assessment: evaluating a knowledge architecture for natural language processing, *Procs. Expert Systems '94*.
- Benford, S., E. Burke, E. Foxley, C. Higgins and A.M. Zin (1994) Ceilidh: a software support system for the automated assessment and administration of computer programming courses, *East West Journal on Computers in Education*, 1(1), 57-68.
- Bowden, P., P. Halstead and T.G. Rose (1996) Extracting conceptual knowledge from text using explicit relation markers, *Lecture Notes in Artificial Intelligence*, 1076, Springer Verlag.
- DTI (1996) *Report of the Multimedia Industry Advisory Group*, UK Department of Trade and Industry.
- Foremski, T. (1994) Overloaded on the Internet, *Financial Times*, 8 November 1994.
- Heppell, S. (1996) Strategy in a world on the move, *Times Educational Supplement IT Review*, 5 January 1996.
- HMSO (1995) *Superhighways for Education — Consultation Paper on Broadband Communications*, Department for Education and Employment.
- HMSO (1995) *Superhighways for Education — The Way Forward*, Department for Education and Employment.
- Hutchison, C. (1996) Snares in the charmed circle, *Times Higher Educational Supplement*, March 1996.
- NCET (1995) *Highways for Learning — The Internet for Schools and Colleges*, National Council for Educational Technology.
- Parker, T. (1994) The Internet and schools: a survey of networking activities, *Proc. of Internet Society's INET'94*, Prague.
- Pickering, J. (1996) Teaching on the Internet is learning, *Active Learning*, 2.



Integrated information education as standard part of university curriculum in chemistry

Jaroslav Silhanek

Institute of Chemical Technology, Czech Republic

Abstract: *On the basis of long experience in teaching chemical literature courses, an integrated education and training programme was introduced as an obligatory course in chemical information where both classical printed and modern electronic forms of information are covered. This obligatory course forms a standard part of the university chemistry curriculum. An important condition for designing this course was building a special classroom for teaching information technology which is located directly in the Central Library of chemical faculties, and where not only is computer access to local and distant information sources provided but also sets of duplicate printed secondary reference sources like Chemical Abstracts (complete five-year period with appropriate Collective Indexes), Beilstein Gmelin, encyclopaedias and others are available for teaching purposes. The syllabus of the programme is given and initial experience with this form of information education is presented.*

Keywords: Information education, printed and electronic technology, chemical information, syllabus, courses in information

1. Historical overview

Some kind of efficiency in extracting information from resources, handbooks, scientific periodicals and other materials has always been considered an important part of the overall professional prerequisite for scientific work in chemistry. A good knowledge of how to find important information has been valued perhaps more in chemistry than in any other scientific discipline, as can be well-documented by an extensive list of textbooks, monographs, manuals, instruction materials and other aids written on this subject (Ref 1).

But formal education in chemical information, or at least some kind of practice session for undergraduate or graduate students, has been conducted mostly individually, together with other subjects or as a part of laboratory exercises. Students have received the most intensive training in library usage when working in laboratories on a voluntary basis, or during preparation of their diplomas and theses. It was (and often still is) quite common and accepted for instructors and professors to take their students into the library to show them on the spot how to work with indexes and secondary reference resources, and where to find primary periodicals and factual data. A plethora of books on chemical information (or literature) is used mostly by research chemists when they realise that, being productive and efficient researchers, they need more detailed information on chemical information methods. This pattern was followed generally in our chemical faculties with the notable difference that we could not buy most of the manuals and monographs on chemical information published in Western countries. Therefore during this period a few such materials were published in the Czech language and, due to growing demand, in repeated editions (Ref 2). This situation with our chemical faculties is not essentially different to the situation in other countries. Zass stated that the number of chemistry departments offering formal instruction in chemical information retrieval is rather small in Germany (Ref 3).

It became slowly apparent that instead of leaving preparation for information retrieval to the individual it would be better to offer at least some form of formal course framework, e.g. a couple of lectures with subsequent practice sessions and preparation of a short review paper on a given subject to get a credit. Courses on Chemical Literature for which we had appropriate instruction materials had been taught this way for more than 20 years (Ref 4). A strong impact on formal education in this subject came after the indexing system in Chemical Abstracts was restructured and refined in the early 1970s. The separation of General Subject Headings and Chemical Substance Headings with a preselected hierarchical list of approved terms for volume indexes, and the open non-hierarchical terms used as keywords in Keyword Indexes, meant that some old practitioners were not able to locate information in Chemical Abstracts in the same manner as they used to be.

The advent of online technology in the former Czechoslovakia at the end of the 1970s, with a central intrastate database centre accessible by dial-up public telecommunication links, made possible the installation of early CA Online products such as CA Condensates and then CA Search (Ref 5). It was another incentive for the slow but steadily growing inclusion of information on online and other electronic services into the course on Chemical

Literature. Finally we decided it was time for a more radical rearrangement of the course on Chemical Literature into an integrated education and training programme, where both traditional printed and modern electronic forms of information would be presented in a comprehensive manner. Moreover, we suppose that it is also time for such a course to be included as obligatory in the standard university chemistry curriculum.

2. Main chemical information course objectives

We formulated three main objectives for the basic obligatory course on chemical information:

- the classical library with printed sources — books, encyclopaedias, periodicals, tables, handbooks and similar materials — will still be used in the foreseeable future and *students have to learn how libraries and their collections are organised, and how to use them*;
- electronic access to practically all kinds of information will grow rapidly, and *students have to learn how to use such access systematically*, not just accidentally by surfing the Internet or similar techniques. The main information sources for which usage has to be mastered in electronic form are the secondary reference databases such as Chemical Abstracts, Beilstein, Gmelin, Science Citation Index and others.
- for chemistry and many related disciplines *the computer representation of molecular structures is one of the key points of the whole system of chemical information*, and therefore it must be covered by the education and training programme.

In practice the first two objectives are closely interrelated: the third represents a somewhat separate and independent subject. We hope that by formulating these objectives for our course we set up priorities not only for the present situation but so that the hierarchy of objectives will survive developments in information technology for some time.

3. Technical prerequisites and requirements

It is clear that for any modern course on information retrieval and technology the possibility of practical work with computing facilities in a network environment is a fundamental prerequisite. For our programme we were able to build a lecture room designed specially for teaching and training of information technology. But what is perhaps more important is that the lecture room is located in our technical university's Central Library of chemical faculties: in other words, it is in a library environment. The natural interrelation between classical printed information sources and modern computer technologies may be displayed very clearly with plenty of practical consequences for training, demonstration and hands-on experience. The lecture room is equipped with computer/terminals for 10 or 20 students, with one computer with an LDC panel for the lecturer or instructor, all connected to the university network in a TCP/IP environment and through it to the Internet.

The interrelation of modern information technology with classical printed services and sources is further strengthened by providing this lecture room with spare sets of the most important printed information sources, like a full five-year set of the printed Chemical Abstracts volumes together with the set of Collective Indexes for the same period. This material comes from a cancelled subscription to CA in reorganised state information centres: we realised its enormous value for hands-on demonstration and usage training without disturbing the regular users of the printed Chemical Abstracts in the reading room of the Central Library. Besides Chemical Abstracts it was not too difficult to find duplicate volumes of other important printed sources — for example individual volumes of Beilstein's Handbuch, Gmelin's Handbuch, some encyclopaedias, older editions of some tables, and many other materials. Each student of the group can thus work with his or her own exemplar and all problems and tricks can be demonstrated on a real information source.

Summarising the technical requirements from the other point of view we can identify the following particular items:

- each terminal in the lecture room is an independent PC, with a set of standard software making it possible to train students not only to communicate with a database but to finalise results into the reports or tables, or to evaluate them statistically by means of other standard software;
- full connectivity to all local sources as well as to the Internet and database vendors, primarily to STN International;
- local Unix servers with loaded subscribed and licensed databases (e.g. CrossFire);
- a CD-ROM server on the local network providing access to databases or other materials on this medium (at present this uses the UltraNet CD-ROM networking technology based on copying the CD-ROM image onto a fast hard disk).

With this local and distant access technology available in the classroom, all possible methods of information retrieval may be taught, demonstrated and practised.

4. Database and software selection

When we are dealing with printed information sources there is practically no problem in selecting one particular product in preference over another. As in most chemical libraries in the world, the printed volumes of Chemical Abstracts with the appropriate indexes and the collection of primary journals represent the standard set of information resources used for both research and for teaching. The situation with electronic resources is much more complicated. Mainly for budgetary reasons it is not possible to have the full spectrum of important sources in both printed and electronic forms, or to buy most of the marketed databases or information system products even if they are very interesting and important. Beside the total price of a given product the most important factor for us is the possibility of using it without payment per use. In other words, we strongly prefer products which may be subscribed to and accessed as a local database on CD-ROM, or delivered on tape, or which may be accessed without limitation for a flat-fee subscription. Besides the self-evident reasons for such a policy, the still existing complication with cross-border payments of relatively small amounts, and the burden of often enormous bank charges, makes even cheap educational access policies such as the KR Dialog/Data-Star Classroom Instruction Programme (CIP) rather unattractive.

Being more specific, from this point of view probably the most important online information source for chemistry, STN International with CA databases, is difficult to teach effectively because STN does not have a flat-fee policy payment scheme. Therefore we can use only the STN Mentor product for online access demonstration and search simulation. On the other hand we are using the 12th Collective Index on CD-ROM extensively in our programme, and direct comparison of printed Collective Indexes with their electronic counterpart (even though it is not the same Collective Index) during classroom lessons is very helpful in understanding the indexing philosophy and rules. But we feel that the extensive database resources of Chemical Abstracts are not exploited for chemical information education and training as they should be.

The situation is quite the opposite with the factual databases from Beilstein and — though this does not yet apply in our case — Gmelin. The integrated package of both Beilstein's printed volumes with unlimited access to the electronic CrossFire version mounted on the local Unix server represents a very attractive solution for teaching not only information retrieval but organic chemistry itself. In our basic obligatory course on chemical information we teach CrossFire only on a very basic level, demonstrating predominantly the structure and substructure searches, but we found it to be an extremely motivating lesson for undergraduates. From this point of view the upgraded version, the CrossFire Reaction Plus which has the ability to search for reactions and to map starting material on products, would be still more attractive. As CrossFire has only been available for little more than a year on our local network the full potential is not yet exploited, and we expect its much wider involvement in different courses both on organic chemistry and chemical information methods. We expect the same development with the electronic version of the Gmelin database, which it has been announced will be mounted in the same CrossFire environment.

As we included not only the passive retrieval of information but also the active preparation or work-up of information in our programme, we have to select appropriate software products for such purposes. From the plentiful chemical text editors we rely on the standard MS Word 6.0, with ISISDraw for the preparation of structures or schemes for export into Word documents. In our programme we used the whole ISIS system for demonstrating reaction searches and for the design and development of structure databases, and we are negotiating with MDL for setting up some form of classroom access to the ISIS modules.

5. Language problems

For non-English speaking countries there is one problem sometimes overlooked concerning the language which is used for demonstrating information retrieval or the databases themselves. A not unimportant part of our programme is that early in the curriculum students are exposed to the fact that the vast majority of information, which is evidently indispensable in any chemistry-related professional work, is in English — or, to a lesser extent, in some other widely-used language. Our course is taught in the students' mother language, but for demonstration and when working with full or demo versions of databases or information systems, we pay no attention to the fact that the material is in English: we leave it to students to cope with this reality. This policy was adopted for a long time with hardly any complaints: on the other hand, the students often appreciate being confronted with the professional language reality early on.

6. Syllabus of integrated course in chemical information

The programme was designed for the present time allocation, which is two hours per week for one term (one semester, i.e. 14 weeks), which makes 28 hours altogether. This course is taken mostly in the second year of study, in the summer (fourth) semester. According to our lengthy experience this is the first suitable term. In earlier semesters students mostly do not realise the importance and impact of professional information, or they have only a superficial or one-sided knowledge of chemistry itself. There are four main blocks which correspond to the main objectives formulated above:

General subjects:

- [1] introduction, basic concepts and elements of information processing and systems;
- [2] the document as the most important form of scientific information;
- [3] information technology, recording and storage systems;
- [4] computer networks, local situation and global system, the Internet.

Chemical information, secondary reference sources, Chemical Abstracts:

- [5] chemical information, history, specific features development and present status;
- [6] basic secondary sources, the 'Handbuch' concept, Beilstein, Gmelin;
- [7] abstracting and indexing services, Chemical Abstracts Service;
- [8] primary document analysis, CA indexes system and their usage;
- [9] comparison of printed and electronic indexes, online searching;
- [10] system of CA databases, methodology and strategy of information retrieval;
- [11] other databases, Science Citation Index, SDI and Current Contents approach;

Computer representation of structure, structure and reaction database:

- [12] chemical nomenclature and computer representation of chemical structure;
- [13] structure databases, exact and substructure search, reaction databases;
- [14] 2D and 3D representation, ISIS system and database combination;

Additional subjects:

- [15] patent information, basic facts about patents and patent systems;
- [16] work-up of information data, evaluation, statistics and presentation.

It may be stated that at this moment roughly 50% of material is treated as printed sources and 50% as electronic sources, with a slow but steady shift to the electronic part. Credit is given on the basis of written tests with individual retrieval problems and a short review paper on a given simple subject. This form of programme has been in development for a couple of years: the final version presented has been tested with more than 600 students with quite favourable acceptance. At this moment there is no doubt about the integration of this course as obligatory for all chemistry students at our university.

7. Commentary on syllabus

It is necessary to add a few notes as commentary on the syllabus presented above. First of all we found the success of such a course depended on starting with very basic ideas, concepts and elements of information processing and information systems. Regardless of the fact that the modern world is flooded with information, students mostly do not realise how the information is formed, processed, stored, retrieved, disseminated, evaluated and finally used. It is very important to demonstrate clearly the cycle of scientific information, going from the known facts to new discoveries, and marking this route by correct citation praxis. Most students for the first time have the typical research paper in hand which should be their companion for their whole professional careers.

In the second block of subjects we concentrate on the main part of the system of scientific information, i.e. on the abstracting and indexing services and their products — the secondary reference sources regardless of whether they are used via printed volumes, or via online or CD-ROM media. Fortunately the Chemical Abstracts Service represents an ideal model for demonstrating and training in practically all the areas of such a business. From the other point of view, to teach students how to use Chemical Abstracts effectively in everyday work and in most practical situations without problems is without doubt one of the most important goals of this course. For this reason the time reserved for practising Chemical Abstracts in both printed and electronic form is greater than would follow from the syllabus. CA indexes are also very appropriate tools for introducing online search methodology and praxis, with the demonstration of basic steps from commands for login, through expand and search, to display. Unfortunately, due to the lack of some form of flat-fee payment scheme we can only use the Mentor module or a live demonstration by an instructor, without the possibility of students working online by themselves. It is rather a pity for we found that the instruction material for using CAS Database on STN covers the subject in an excellent manner and we are using it in many instances (Ref 6).

In this block there is also the common problem of going from citation in a secondary source to the location of primary information in its original form. Students are forced to decipher the full name of the primary source from its abbreviation, for which the CASSI Index is available and strongly recommended, and to find the original paper on the library shelves. Some mistakes introduced deliberately into the citation make an interesting exercise of the task, demonstrating the need to pay great attention to the correct citation.

The next most important subject is the computer representation of chemical compounds structure, starting

from connectivity tables and going through drawing software and structure databases to the sophisticated reaction databases, in the last instance at the demonstration level only. Here the marketing strategy of the Beilstein Institute for its CrossFire product, based on a flat-fee budgeting policy, very substantially enhanced the possibility for practice sessions in structure manipulation. This part of the programme is both very attractive for the students and also helps in understanding some structure, reactions and in the case of the CrossFire database factual relations from general organic chemistry (e.g. to correlate structure with melting points or other properties of molecules). We suppose that still greater attention will be given to this part as some structure and reaction databases are now practically in the category of industry standard software.

As an additional subject we devote at least two hours to patent information, where the understanding of general problems is rather poor, and some time is also spent on the final work-up of data found using the standard software system.

8. Conclusion

While the idea of integrating classical printed and electronic forms of information into one formal course in chemical information has been developing slowly for some time, a more radical rearrangement of lectures and practice sessions could be realised after building the special classroom for teaching information technology inside the Central Library at our university. With such technical background and support available we had far greater possibilities for selecting course modules. Both students and instructors greatly appreciate the possibility to work, explain and discuss problems with authentic volumes and indexes of Chemical Abstracts in a classroom environment without disturbing other library visitors. Carrying such a course practically inside the library also offers many possibilities for demonstrating and training in many practical problems connected with real information retrieval tasks.

However, the practice sessions on computer and electronic information methods may also profit from immediate access to standard library resources, and can provide students with a good and reasonable feeling for the adequate selection and efficient usage of information resources and technologies. Summarising these favourable factors, we could introduce a formal course on chemical information as an obligatory course for all students in a basic three-year basic course of study in our chemical faculties. Until now the programme has been tested with more than 600 students who took this course mostly in the second year of study. The acceptance was quite favourable and the course is now a standard part of the university chemistry curriculum.

Jaroslav Silhanek
Institute of Chemical Technology, Prague
Technicka 5
166 28 Praha 6
Czech Republic
Tel: +42 2 311 50 81
Fax: +42 2 311 96 57
E-mail: silhanek@vscht.cz

References

- [1] Warr, W. and C. Suhr (1992) *Chemical Information Management*, VCH Weinheim, and references cited there.
- [2] Hanc, O., B. Hlavica, V. Hummel and J. Jelinek (1961) *Chemická literatura a její využití v praxi*, 2 ed., STNL Praha.
- [3] Zass, E. (1995) German databases in the chemistry curriculum at the ETH Zurich, in Kristina Voigt *et al.*, German databases for national and international users, using environmental databases as an example, *Online & CDROM Review*, **19**(2), 75-77.
- [4] Silhanek, J. (1983) *Chemická literatura*, 3.ed., STNL Praha.
- [5] Silhanek, J., P. Rosendorf and J. Orsag (1994) The recent development of the academic and commercial network services in the Czech Republic, *Proceedings of the 18th International Online Information Meeting*, 6-8 December 1994, London, Learned Information Ltd., Oxford, pp. 261-267.
- [6] *Using CAS Databases on STN*, Student Manual (1995) CAS, January 1995.



An iterative development of a training programme in the use of the Internet as an information source and publishing mechanism

Gobnait O'Riordan

University of Limerick, Ireland

Abstract: A team of library and computing staff was used to develop an Internet training programme at the University of Limerick. Its objectives were to present the Internet as both an information source and publishing medium, and to train faculty and staff in the effective use of these applications. The team adopted an iterative approach to the programme by incorporating review and revisions at the end of each training cycle.

Keywords: Internet, training, information source, publishing

1. Introduction

The purpose of this paper is to assess the development of an Internet training programme for the staff and faculty at the University of Limerick. During an eighteen month period from the Spring of 1995 a three-cycle programme concentrated on raising awareness and developing proficiency in using the Internet as a source of information, and on training end-users to become independent information publishers on the university's Web site. The programme evolved as a response to these needs and as a result of a general demand for information about, and access to, the Internet by staff and faculty.

The University of Limerick has a student enrolment of 8000 and employs 642 people. The campus is fully networked and provides access to information sources, and communication and document creation software. The Information Technology Department (ITD) and Library and Information Services (LIS) are two units of the Information Systems and Services Division (ISSD) with responsibility for Internet services. Prior to the development of this training programme, staff from both areas were involved in two related projects. The earlier was the design and development of ULIX (University of Limerick Information Exchange), a Web-based information exchange or publishing mechanism known as a CWIS or Community Wide information System. These systems are defined as 'a means of disseminating information electronically to the people within a single institution' (UCISA 1995). It is primarily a communication mechanism that acts as an access point to documents available via the World Wide Web, and as a means of publishing by and to the local community and to the world. It also serves as an Internet on-ramp for faculty, staff and students. The second project, EDUCATE (End-User Courses in Information Access through Communication Technology) was a CEC Libraries Programme project to produce a self-paced user education course in the selection and use of information sources. The project had a particular focus to deal with online information sources, and included real time online demonstrations and exercises.

Both projects were organised on a cross disciplinary team basis and included staff from ITD and LIS departments. They were of benefit to the development of the Internet training programme as they provided an operational model, experience of IT and library staff working together, and knowledge of the Internet. It is also significant that they established a role for librarians in the development of Internet services at the university.

2. Research methods

We were unsure as to the best method for delivering such a training programme, as the technology was developing at a rapid pace and we were in many cases still trying to come to terms with it ourselves. The literature offered little guidance as to the effective delivery of Internet training in general or specifically with regard to an academic community. We decided to adopt an action research methodology, a style of research common in the social sciences and classroom-based research. This method advocates diagnosing a problem and trying small scale interventions in the context of a specific real world situation. Such modifications or interventions are constantly monitored within the ongoing situation with the objective of improving practice.

3. Training methods

The team approach ensured that all courses were offered jointly by the training section of ITD, the College Librarians of LIS and staff of the ULiX office. Each course had at least two presenters, one from each department, and a minimum of three demonstrators assisting clients as needed. From the outset the courses were designed to maximise client participation with short demonstrations interspersed with hands on and exercise sessions. Courses were scheduled for three hours with a half hour break for tea or coffee. This break provided a useful opportunity to assess client response and to deal with individual difficulties. We limited class size to sixteen, one person per PC. Where possible we used the technology to deliver the training utilising MS PowerPoint presentations and purposely designed pages with live links projected onto a 30" monitor. Such an approach was open to system failures but we sought to use these occurrences to teach clients how to deal with the problems encountered in accessing the Internet. The courses were presented in a casual and informal manner so as to put participants at their ease, as many were uncomfortable appearing as novices. Courses were advertised widely throughout the university using print and electronic media. The logistics of bookings and scheduling were handled by the training section.

4. Training programmes

Our objectives in offering an Internet training programme were threefold. At the outset we aimed to raise the level of Internet awareness among the university community, specifically of its academic applications. Having established a reasonable level of awareness, our more specific targets of utilising the Internet as an information source and as a publishing mechanism were pursued. To deliver such objectives three courses were required. An introductory session, in the initial cycle, was quite informal with participants free to drop in at any stage during the session for a cup of coffee, followed by a one-to-one guide to Netscape and selected resources, and an opportunity to stay and browse at their own pace. An intermediate course covered information searching and more advanced Internet features such as FTP and Telnet. An advanced course dealt with home page creation covering information structuring, style guide and basic HTML.

Of the three training cycles conducted so far, the first cycle concentrated on introductory cafe type sessions. The second and third cycles continued to offer this but our focus was on the development of intermediate and advanced courses. Each cycle was followed by team review and revision where necessary. During the development of the programmes we gradually moved from a tool-based or systems training approach to a more user-focused one. This was due in part to developments in the technology, which saw a more integrated approach using one browser to access many applications, and to a greater emphasis on use or application aspects in the courses. The multidisciplinary team approach to delivery of Internet training brought together these two approaches. The ITD staff had valuable experience in training in and use of computers, and an extensive range of software packages. Librarians had a greater user and usage contribution and brought knowledge of the information searching practices, user needs and the value of such applications and sources to the university community. We faced the challenge of integrating both of these approaches into one unit during the three cycles of programme development.

5. Facilities and documentation

All courses were held in a purposely designed PC teaching lab and access to such a facility greatly facilitated delivery of the programme. However we were in competition with undergraduates for access to the room and access times were not sufficient to meet our needs.

We prepared paper-based documentation to support the courses in the initial programme cycle. These comprised an IT manual, subject-based site lists and style guides for document creation. These took up a considerable amount of time to prepare and very soon needed revision as a result of technical advances and changes in course content. We did not have the time to undertake such work and were conscious that we were in many cases recreating material already available on the Internet. Instead we moved to creating purposely designed Web-based training pages that would incorporate our course structure with useful Web sites and lists. This work is still under construction and will, when completed, be retained as a permanent feature of ULiX.

6. Programme assessment and redesign

Each training cycle finished with a review meeting in which all involved were encouraged to review the courses so as to improve the next cycle. The introductory cafes of cycle one were considered too informal and lacking in content. We were emphasising the idea that the Internet was a fun concept and the jump from this to the serious applications in the intermediate course was too great.

It was replaced in cycle two with a more formal three hour course that provided a detailed presentation of Netscape features and demonstrations of FTP and Telnet. Again ample time was given for participants to spend time browsing. Our objective had been to place greater emphasis on the value of the Internet to the university

community: however the formality and expanded content were too complex for beginners. Following the second cycle we made further revisions, making it shorter and less formal but retaining as much as content as users could handle. The changes made in the introductory course had a direct effect on the intermediate course, decreasing or increasing the amount of coverage given to FTP, Telnet, customising Netscape and so on. In the first cycle a second theme of the intermediate course was the use of the Internet as an information source and this aspect grew in significance in subsequent cycles. Initially we sought to give this a strong subject-based orientation by offering subject-based sessions, e.g. Internet for Engineering, on topics corresponding to the main subject areas of the university.

This approach was not as successful as we had hoped. It was difficult to get academics scheduled into appropriate subject groups, and there was often such a spread of topics and sources within each subject that the description and assessment of many sources were irrelevant to much of the class. Although we spent a great deal of time assessing sites, and creating annotated subject lists and transferring this knowledge, the participants did not build up real search skills for themselves. We revised this course to one that offered a generic approach to information searching. We concentrated on how information was organised on the Internet; what could be expected from it; how best to use subject catalogues, discussion groups and search engines; finding people; and how to keep up to date. Following these revisions participants had a greater awareness of the value and pitfalls of the Internet as an information source, and how best to access it.

The advanced course required least revision, with the addition of such features as how to publicise your pages, a greater range of page templates, and images and photographs. We simplified the procedure for FTPing locally created pages on to the server and directing participants to HTML verification services. This course will require some modification as we intend to make an updated editor available shortly. In addition we now have a demand for a more detailed course dealing with features like Java, tables and sound. As the demand for information on many of these advanced aspects of Web publishing comes from individuals we are considering offering a clinic or question and answer style session with a panel of 'experts'.

7. Training issues: the Internet as an information source

The design and organisation of these courses raised specific issues relating to training in using the Internet as an information source. As with training in using other electronic information sources we focused on search strategies, Boolean logic and search engines. It was not enough to teach how the Internet works: in addition we needed to develop an understanding of the nature of information on the Web. There is so much hype and publicity about the vast information superhighway that it is important to help participants to understand what it is and is not. One technique we used was to contrast the reality of the Web as an information source with the media image. It was important to assist participants to discriminate between information resources, to understand what is being searched, how various search engines work and what levels of quality can be expected, and to warn of the instability of many sites. Another technique used was to compare it with other information sources, both print and electronic. The Internet offers additional information source to those already available, and it was important to remind people of these and assist them to choose the source most appropriate to each requirement. In addition the Internet in its own right offers access to a vast range of new information products and services. We were surprised at the extent to which we needed to teach critical thinking skills in using the Web as an information source. These greater choices and options demand more effective search skills, search software knowledge and source evaluation techniques than ever before. Yet all of this is presented in an easy-to-use and sophisticated mode that belies the complexity and demands of the medium. The general image of the Internet presented in the media is of an easy to use medium, and participants often wondered why it was not. Ease of use was strongly correlated with previous familiarity with Windows-based software.

The change from a subject-based to a more transferable search skills approach to training in use of the Internet as an information source was taken in part as a response to local conditions. We still perceive a need to incorporate a subject-based element in such a training programme, perhaps as an additional course following the generic searching course. There is a need to provide greater linkages with other information skills training courses, and to put more effort into presenting the Internet as an additional information source and as a means of accessing local and remote databases.

8. Training issues: Internet as a publishing mechanism

From the outset we sought to use the training programme to develop a cohort of information providers who would in time manage various pages on the university's official server ULiX. At each level of training we promoted the Internet as both an information source and a publishing medium, and that participants could be both users and publishers of information. We would facilitate either or both options: the choice was with each individual.

Initially participants were most concerned with the apparent difficulty in mastering HTML and FTP. The advanced course covered these basic skills, and with regular support and back up from the ULiX office many gained an acceptable proficiency. It was more difficult to get participants to focus on the issues of information management, although these in the long run proved more challenging. We emphasised the need to put a considerable effort into how they wished the information to be used, to focus on the needs of their potential readers, to adopt good quality style and design standards, and to update and publicise their pages. Thus in addition to the

technical skills needed to create Web pages, information providers needed basic skills in Web writing, editing and publishing. Many participants were enthusiastic to 'put up a page' and often lost interest at that stage. It was a challenge to promote the benefits of the Internet as a publishing medium to carry people past this stage.

The Internet offers extensive opportunities for the academic community to publish a wide variety of information both to an internal and external market. Participants were slow to identify many of these applications. We made a conscious effort to present specific examples suitable to this community, initially from external universities and later using internal examples. ULiX is now used to publish a wide range of documents including courseware, policy documents, minutes, newsletters and research project material.

At this stage we are encountering legal and policy issues, and have drafted policy statements governing electronic publishing and information providers for adoption by the university. We are also encountering a growing demand for training in advanced features of Web publishing. To provide training for these aspects and to sustain continued use of the Internet as a publishing mechanism, we are planning a series of seminars to support and enhance the contribution of local information providers.

9. Course participants

All participants were self-selecting and attended the courses out of interest, often as a result of publicity on TV, in newspapers and magazines, or as a result of general presentations on the ULiX service. They were keen to try out the medium themselves to experience 'surfing the net'. A minority had grasped the implications of the Internet as an information source and publishing mechanism, and the applications it could have for themselves, their departments, professional associations and research groups. This understanding grew following attendance at the introductory courses or ULiX presentations and provided the market for the intermediate and advanced courses. It was significant that participants wanted to build up skills in Internet searching and publishing themselves rather than use an intermediary. Demand for these courses was high: over 300 initial bookings were made and 175 participants eventually attended an Internet training course. These were drawn in equal proportions from faculty and staff, and from both technical and non-technical areas. Indeed there was a strong demand from Informatics and Computing faculty for intermediate and advanced courses. Initially demand was greatest for introductory courses but this has waned somewhat as demand for the intermediate and advanced courses continues to rise. Results of a survey of participants will be available at the conference.

10. The trainers

The trainers for these courses were drawn from either a computing or library background. The computer training staff had extensive experience in end-user training in a wide variety of personal productivity tools, and would have worked with many of the participants previously. They were comfortable with their training role and in using technology to deliver training. The librarians also had experience in providing end-user training and in delivering information skills courses. However these activities only comprised part of their work and were concentrated into short periods of time. Although most had previous teaching experience at school or university level, all described themselves as uncomfortable with the training role. In particular they were uneasy in working with groups, preferring to work on a one-to-one basis with clients. In addition they were not skilled in using technology to present their segments of the course.

All trainers became involved as a result of a convergence of interests. ULiX staff required such courses to generate awareness of the service and to train self-sufficient information providers. Computer training staff were aware of a growing demand for some Internet training from faculty and staff. The librarians were exploring the Internet as an information source in subject-based groups with colleagues from other Irish universities in preparation to demonstrating it as an information source. However the programme was started from the ground up on a trial basis by individuals working in a voluntary team, as we realised that we could more effectively and easily develop such courses together. Senior management was not involved in the programme either at initiation or development, and work in this area was never specifically referred to or acknowledged.

All the trainers indicated that they benefited from the working on the programme. The computer training staff and college librarians developed a greater awareness of ULiX as an access point for Internet resources. Computer training staff did not find it difficult to develop the skills needed in using Internet applications, and found it less difficult and time consuming than many other software packages. Indeed during the course of the programme Internet access became standardised into one browser and that task became even simpler. They learned to incorporate a greater use and user orientation into the course rather than a systems approach. The librarians had to build up computing and Internet skills in addition to developing some knowledge of the resources and the search tools available. Assessing a growing number of search engines, how to use them and what exactly each one searched was an important issue. The very growth in the range and quality of the resources becoming available on the Internet during the period provided an additional burden. The librarians also developed skills in MS PowerPoint and using specially designed Web pages to deliver course information. This programme forced them to learn the skills needed in other areas of their jobs to demonstrate effectively the range of electronic products being made available on campus during this period. The computing training staff did not spend a lot of time in preparation, and courses could be rerun without much rechecking of sites, but they were involved in every course so delivery time was high. The librarian's involvement was concentrated in the interme-

diate section, where the focus was on searching and a great deal of time was needed in preparation, especially in relation to the time spent on delivery. The ULiX staff were involved in all courses, either making presentations on the service or presenting the information management and style guidelines elements of the advanced course. In addition they assisted in course and review organisation, and provided support for the practical hands-on aspects of all courses.

11. Team training

We used a team training structure to deliver these courses, due in part to local factors rather than in any belief that it was the best method. The ISSD structure, based on a converged IT and library model, had prior to the Internet projects little real convergence apart from senior management. The use of a team approach to develop the university's Web site was continued to its presentation to the community and to the training programme itself. There was a realisation that together we could offer a more effective and useful programme than competing with each other. Such a team training approach was beneficial in that it provided an opportunity to create new working and training relationships based on shared knowledge and professional resources.

Team management was informal and based on the voluntary participation of all members. Group meetings were augmented with one-to-one meetings used to encourage members, identify difficulties and suggest solutions. In these regular review sessions individuals had an opportunity to learn from each other, and to share their views on content and training methods. Many areas of benefit emerged and information was exchanged regarding equipment, training skills, information sources and search skills. This forced all of us to acknowledge contributions and views other than our own. Indeed, bringing different views together often led to greater analysis and adoption of new solutions. The iterative process was liberating in that we could try out different approaches without any group or individual becoming possessive or defensive. It gave individuals and the team an environment in which to consider and adopt changes and improvements in a natural and non-threatening manner. In this way clients received a more complete and valuable programme, and trainers were able to develop individual skills and a greater appreciation of the range of contributions needed.

The team approach was not without its disadvantages. There were considerable differences in skills, attitudes, motivation and working style between individuals, and in particular between the computing and library elements of the team. There was also the now almost traditional power imbalance between the technical and information management approach to contend with. There was a lack of trust and openness between the two groups, and differences in approach and methods were often not articulated or confronted by the group. Initially the team experience may have reinforced stereotypes and hindered co-operation. Although working together has resolved some of these difficulties we still need greater clarity, understanding and acceptance of the knowledge, and the contribution of each group. Working in a team without the more familiar authoritative decision making mechanisms required skills unfamiliar to the team participants. We need to develop improved team skills so as to foster greater role negotiation among members.

12. Implications for information professionals.

Much has been written on the implications of the growth in electronic information for the role and indeed the very future of the information professional. Internet and Intranet development towards a standard interface for accessing a vast range of resources, legacy and new databases, and communications mechanisms is a significant milestone in that debate. Our experience in presenting the Internet as a publishing mechanism and an information source provides some pointers for the evolving role of the information professional.

The greatest difficulty among information professionals is one of role perception and attitude. There may not be a great awareness outside of the profession of the contribution we have to make: however, this is not an insurmountable problem. Though it required some effort we were successful in securing an equal role with our computing colleagues in the Internet training programme, and course participants accepted our role without surprise or difficulty. I believe far greater difficulties lie in our own attitudes and in a serious lack of confidence in the value of our contribution. Librarians tended to accept the technology in a submissive manner and seem unable to challenge the IT professionals on equal terms. We need to realise that information management, and organisation and information search skills and manipulation, are the crucial issues facing the electronic information world; not the technology. The technology is getting easier to use and is solving its own problems while the information problems are growing at a rapid rate. Clearly the IT focus is one of presenting tools and applications rather than use and value. It is up to the information professionals to realise this, and to present and deliver an alternative approach to electronic information creation, selection, acquisition and use. Once the initial technical mystery has dissipated, clients quickly grasp the value and need for information management input and appear to have no reservations as to our role in this.

We cannot assume that our traditional clients will turn to the library for such support. The information market place is an increasingly crowded one and technology facilitates direct end-user access. Our training courses should be developed in line with best educational practice and validated as appropriate. Our services need to be targeted to satisfy user requirements, both perceived and unrealised and delivered to commercial standards. As a matter of urgency we must develop or obtain access to promotion and marketing skills, and actively market our services in professional business like manner.

With the emergence of a greater range of resources, multiple access points and greater variety in pricing and delivery options, our clients require more training and support than previously. However our clients, both corporate and individual, need support and training as information providers and publishers, not just as consumers. We need to train and support them in identifying the full spectrum of their information needs and in choosing the routes most suited to resolving them.

It is essential that information professionals develop additional skills as a matter of urgency. We need to develop high levels of competency in standard information creation and manipulation software packages, so that we have a confidence in dealing with the technologies and the technologists. We need to acquire professional presentation and training skills to the highest level. We need to stay abreast of current developments in the IT and electronic information areas as a matter of priority.

These new services and skills must not be viewed as additional extras to our existing workloads, but as essential requirements to move our library services into the emerging electronic information era. We cannot afford to wait for new generations of information professionals to emerge, but must train and retrain now and integrate such skills and services into our daily working lives.

13. Conclusion

All team members saw the future in terms of considerable change for Internet training and in a broader sense. Clearly the Internet technology is developing itself and in the longer term there will be issues concerning the use of the Web to support clients in terms of online training and one-to-one support. As the Internet emerges as a central access point, we need to focus on how we will integrate it with other information sources, both electronic and paper-based, and on how this will impact on end-user training and library services.

In the short term we are considering the development of more modular skills-based units with greater use of support material available online and in print. There is some divergence among the trainers as to whether this could or should be targeted on a subject basis, or whether the generalist approach should be retained. As more resources become available in this medium a subject focus may be necessary in some of the modules. It is likely that over time the Internet will become an element in a broader information skills programme which may be rooted in the general student curriculum, as well as being available to new staff and faculty. We intend to make greater use of the ULiX structure to create programmes and modules, to provide documentation and to gain feedback from participants.

The Internet training programme at the University of Limerick was a useful process in many aspects. That it was an iterative or action research-based approach allowed us the freedom to develop the programme on a trial and error basis, and in so doing to develop new skills. Working in a team training environment facilitated cross-discipline working relationships and knowledge transfer. It was the first step in dealing with an information future that we know will change and develop rapidly in the short term. The programme has however given us a mechanism and a direction so that we can adapt to those changes. The programme was beneficial in that awareness among our clients of the Internet, the Web and the ULiX service has grown. Faculty are becoming aware of the resources available and of the skills needed to identify and access them, and have begun to request similar courses for students and visiting groups. The university now has a functioning cohort of information providers who are using the Internet to publish information to the campus and to the world.

Acknowledgements

I wish to thank those involved in developing and delivering the Internet training programme for their contribution to this paper. They are: Mary Dundon, Patricia O'Donnell, Clare McCracken, Paul Murphy, Pattie Punch and Mairead Waters.

Gobnait O'Riordan
Information Systems and Services Division
University of Limerick
Limerick
Ireland
Tel: +353 61 202193
Fax: +353 61 338044
E-mail: gobnait.oriordan@ul.ie

References

- Cohen, L. and L. Manion (1994) *Research Methods in Education*, Routledge, London.
- Hert, C. (1994) A learning organization perspective on training: critical success factors for internet implementation, *Internet Research*, 4(3), 36-44.
- Makulowich, J. (1995) Training the trainers, *On The Internet*, March, 12.

- O'Riordan, G. (1995) Internet publishing case study: ULiX and EDUCATE, *IATUL 1995 Conference Proceedings*, University of Twente, Netherlands.
- Schiller, N. (1994) Internet training and support academic libraries and computer centres: who's doing what?, *Internet Research*, 4(2), 35-47.
- UCISA (1995) *CWIS Managers Handbook*, Universities and Colleges Information Systems Association, Great Britain.



Distance teaching in online searching

Ina Fourie

Department of Information Science, University of South Africa, South Africa

Dorette Snyman

Library Services, University of South Africa, South Africa

Abstract: Information specialists of the future will have to improve their online searching as well as teaching skills. The latter are necessary to train the growing number of end-users. Distance teaching of online searching poses a number of problems especially with regard to its practical aspects. The principles of sound instructional design and technological developments can, however, be used to develop a multimedia study package that can solve these problems and be used to teach end-users, information specialists as well as Information Science students. At the University of South Africa (Unisa), the Department of Information Science uses a multimedia study package to teach online searching and to make students aware of teaching principles. The Library Service offers online training courses for their end-users. Combining the experience and expertise of the Department and of the Library, the objective is to develop a multimedia study package that allows for independent study where the student need attend only the practical sessions. The application of instructional design to the study package is explained and examples are given of the content, objectives, learning events, teaching methods and media combinations. The study package is suitable for different levels of online searching and for people from both developed and under-developed environments. The interest of adult learners is also catered for. The proposed package includes the use of printed media, videos, practical exercises, workbooks, computer aided instruction, workshops, e-mail and use of the Internet. The package is based on the characteristics of distance teaching and multimedia study packages.

Keywords: distance teaching, end-user training, Internet, online searching, multimedia study packages, South Africa

1. Introduction

The working environment of information specialists has changed and with it the kinds of tasks they perform. Today they must be able to conduct complicated online searches, and design and evaluate databases and database programmes. Furthermore they need to use the Internet, keep track of new developments in the online industry and train the growing number of end-users to use the wide variety of information sources and services available. Information specialists of the future should therefore not only master the more difficult skills of online searching, they should also develop teaching skills. A positive attitude and adaptability to their ever-changing working environment is important.

Teaching skills are not covered in most library science programmes. The Department of Information Science at the University of South Africa (Unisa), however, developed a multimedia study package for the distance teaching of online searching to undergraduate students in which teaching skills are emphasised. The package is designed to address the more difficult aspects of online searching and to stimulate students' interest in the online industry.

The Unisa Library has identified the need to train their growing number of end-users. At present the library presents workshops in online searching to academic, research and administrative staff. Most of the Library's student users, however, study by means of distance education and seldom visit the main library. Training programmes for these users need therefore to be offered by means of distance teaching.

Since the Department of Information Science is already using a multimedia study package it was decided to unite efforts in order to develop a programme to be used within the context of distance teaching catering for the needs of people from different backgrounds. The Department's expertise in instructional design and distance teaching and the Library's experience in online searching would complement each other. In order to support cooperation, all the Library staff responsible for end-user training attended courses presented by the Department to refine their searching and teaching skills. Both parties also emphasise the necessity to keep abreast of new developments.

Our point of departure is that distance teaching poses problems in addressing the *practical skills* of online searching. Technological developments and sound instructional design can be used to design study packages to overcome these problems. Since a multimedia study package offers the ideal solution it will be explained how instructional design is applied in the creation of such a package. The package can be adapted to the training of library, academic and research staff, distance users and information science students. Furthermore, it offers a solution to the problems experienced by users from widely differing educational backgrounds.

2. Distance teaching

Distance teaching differs from conventional or face-to-face teaching. It is not based on the attendance of classes, although there are some distance teaching courses that include a number of attendance sessions. The different approaches to distance teaching are discussed in Refs 10, 11 and 13. Distance education is based on the geographic separation between the student and the teacher. There are also other characteristics:

- although the student is geographically separated from the teacher and teaching institution, contact sessions such as workshops or video conferences may occasionally be offered. Such classes should however not cover too much of the study time and should be planned only for those aspects that *cannot be taught in any other way* (e.g. *practical online searches*);
- the teaching institution supports the students through the planning and development of study programmes, and it provides for the evaluation of the student's performance. It also provides guidelines, motivation and other forms of support. The distance library offers similar support in providing access to information sources;
- distance students are mostly (working) adults studying separately from each other. This should be recognised in the teaching approach. The needs, backgrounds, age groups and experience of the students should be catered for (their subject interests in online searching, for example, will differ). Although adults prefer to study independently they also need support and guidance in their studies;
- any technological media can be used, ranging from the printed media to video conferences, computer-assisted instruction and the Internet;
- although distance teaching is based on one-way communication in which the study material is sent to students, there should also be opportunities for the students to communicate with teachers and their fellow students. The latter (two-way communication) can especially be improved by means of interactive technology such as the Internet (Ref 6).

Owing to its characteristics, distance teaching offers benefits especially to working adults, people unable to attend classes, countries where geographical distances influence teaching opportunities and subject areas requiring in-service training. The benefits include overcoming geographical barriers, enabling students to continue with other responsibilities, financial benefits and allowing students to study according to their own time schedule, pace and learning style. These benefits will not be dealt with here since more information can be found in Refs 1, 2, 3, 8 and 9. These benefits should be considered when planning a *true* distance teaching programme, especially when planning for practical sessions.

2.1. Unisa environment

Unisa practises distance teaching according to the characteristics sketched above. Attendance sessions are kept to a minimum and students receive all their study material in the form of study guides and tutorial letters. Sometimes additional media are included: for example, audio cassettes, videos and computer-assisted tutorials.

The Unisa Library renders an information service to approximately 3000 academic, research and administrative staff on campus as well as to approximately 30,000 off-campus or remote students. An extensive and high-quality information service is offered by a team of subject librarians to postgraduate students from honours through to doctoral level. Information searches are conducted on request for these students, using all available sources. Almost all contact is by means of mail, e-mail or telephone and students need not visit the Library. All books, periodical articles and other material can be sent to students by mail. Unisa has the largest academic library in South Africa with 279 staff members. The Library is fully computerised and makes extensive use of local and overseas databases, e-mail and the Internet. It has a computerised catalogue and a campus-wide CD-ROM network with 42 databases available. There are branch libraries in five areas of the country which supply material needed by undergraduate students. Although these branch libraries are connected to the university network, they do not have access to the CD-ROM network or the Internet.

In the light of changes in the online industry, the Library has embarked on programmes of end-user training because it felt it was important to promote these essential life skills. If the problems encountered by distance teaching are overcome, an important contribution can be made to the information skills of the working adult.

Apart from the typical problems of distance teaching (such as studying in isolation, lack of communication, difficulty in attending practical sessions, catering for heterogeneous groups), Unisa has the additional challenge of both students and staff coming from widely differing environments. Some are highly computer and library literate while others have never been into a library or worked with a computer before. Most South Africans do not have access to a computer. In some parts of South Africa there is no electricity, while some of the urban companies and universities have access to the latest technology. The ideal is to present training programmes to

enable people from both ends of the spectrum to master basic as well as sophisticated online searching skills and to help them gain the necessary computer skills required for online searching.

Recently Unisa's policy was adapted to allow for the use of basic (print-based) study packages as well as for the use of advanced technology for those students in a position to make use of it. There is however to be no discrimination between students with regard to opportunities for passing the examinations. Since mid-1996, Unisa also started using the Internet as an *optional* medium for the delivery of study material and the submission of assignments. The Internet option (called Students Online, or SOL) also allows remote access by students to the library catalogue and stimulates interaction between students and lecturers.

Bearing the characteristics of distance teaching in mind, as well as the University's policy on the use of technology, the Teaching Department and the Library decided that a multimedia study package would allow us to train different target groups at Unisa, namely:

- Information Science students;
- library staff;
- academic, research and administrative staff;
- students (undergraduate as well as postgraduate).

A study package for post-graduate students is being developed.

Since we are dealing with training programmes, the term 'students' (and not library users) will be used to refer to members from the different target groups.

The advantage of a multimedia study package is that it can allow the distance student to:

- study independently but with the support of the teaching institution;
- gain the prerequisite knowledge and skills;
- master theoretical aspects underlying the practical work without access to advanced technology;
- study at their own pace;
- study according to their own style of learning;
- evaluate their own progress;
- use a variety of media;
- communicate with other students when necessary;
- complete separate modules according to their own interest.

A multimedia study package can also allow for occasional workshops that the student attends or for other forms of practical work. The idea is however that the student should be well-prepared and that workshops *et cetera* should focus on imparting practical experience in online searching and not on lecturing.

The following developments at Unisa will also support the use of a multimedia study package as more students gain access to computers:

- more study centres are to be built all over the country;
- the Library is planning an electronic classroom on the main campus;
- the possibility of allowing students access to computers at community centres is negotiated.

Although every effort is made to increase students' access to computers (and therefore the opportunity to practise online searching skills), it must be emphasised that the approach we have adopted is based on the recognition that:

- most of our students are computer and library illiterate;
- most of our students do not have access to computers;
- there are different target groups with different levels of interest in online searching;
- the advantages of distance teaching should be borne in mind (i.e. the training programmes should not be based on attendance classes).

Bearing the above-mentioned aspects in mind, we also feel sure that the theoretical principles on which the proposed study package is based will also be relevant for more highly developed countries.

To design a study package catering for the conditions that have been sketched, sound instructional design, the use of available technology and knowledge of the mode of distance teaching proved to be components of the solution.

3. Traditional approaches to online searching

Online searching has been taught since the early 1970s and a lot has been said about the learning content, objectives, possible learning events and the need for practical experience. A variety of methods can supplement practical sessions: computer aided instruction, demonstrations, emulations or simulations. There is, however, still a requirement to *practice with the real product*, like searching a database or developing a database.

Recent developments have also made it necessary to debate the concept of 'online searching' — does it include CD-ROM searching and Internet searching? Our point of departure is that online searching is the generic

concept, including searching through different media, on local or remote databases. Internet searching is included but should be based on the basic principles of online searching. In this approach the Internet is taught alongside other types of online searching.

There is however little evidence of the use of instructional design, including the recognition of the needs of the working adult learner and the use of distance teaching methods in the literature. Our experience at Unisa may therefore contribute to a better understanding of distance teaching in online searching.

In 1995 the Library started upgrading its staff's online searching skills, before presenting workshops to the academic and research staff. The workshops covered searching on Sabinet as well as using the Internet. In 1996 workshops for the same target group in CD-ROM searching were introduced. When developing a study package for other Unisa users, the approach followed for Information Science students will be adopted: that is, viewing the Internet as one of the many types of online searching. It is anticipated that in time the library staff training will follow this approach.

4. Instructional design

The value of instructional design for distance teaching has been well argued elsewhere (Refs 7, 9). Instructional design is regarded as '... the process of specifying conditions of learning. Its purpose is to create strategies and products at the macro level, such as programs and curricula, and at the micro level, such as lessons and modules' (Ref 14). It requires an instructional design model, of which a number have already been suggested in the subject literature for teaching in general, and for distance teaching in particular (Refs 4, 5). An instructional design model provides a set of logical steps which enable one to design training programmes based on the careful analysis of the situation and by applying didactic or teaching principles.

In instructional design, the specific teaching situation is analysed to identify the following:

- *the needs of the target group(s).* At Unisa we have more than one target group, which means that the programme will have to allow for different levels of sophistication. Information Science students and Library staff would be taught to do more advanced searches. They should also be introduced to the teaching principles underlying the design of the programme;
- *the entry level and background of the target group(s).* Do they know how to use a library; have they worked with a computer before; do they have access to a computer, *et cetera*. Depending on their existing skills, opportunities will have to be provided for them to gain prerequisite skills such as computer skills. At Unisa the Information Science students can attend an orientation lecture on computers where they work through a computer-assisted tutorial that includes sections on the use of a keyboard and the basic functioning of computers. Students are also advised to purchase the programme if they have access to a computer. Since practising librarians need to know more about the use of computers (e.g. basic DOS and Windows, and hard disk maintenance) a special course was planned for them. Similar methods will be followed to ensure that the other students will gain computer skills;
- *the technology to which the target group(s) have access.* Library staff and other Unisa staff have access to computers. Using the computer network to deliver the study material or for practising the online skills should therefore not be a problem. Most of the distance students, however, do not have access to a computer. Although the situation is improving, a training programme for students should be based on the use of printed media, (e.g. study guides or workbooks). Workshops should be organised to teach practical skills.

Instructional design involves more than deciding on whether to use printed media, computers or workshops to present the learning content. The importance lies in the logical presentation of the learning content and the use of performance objectives. It is also important to get students actively involved and to allow them to monitor their own progress. These as well as other things are catered for by multimedia study packages.

The crux of the design of a distance teaching programme therefore rests on the following aspects:

- effective use of the possibilities that are offered by the characteristics of distance teaching (e.g. including occasional workshops);
- the application of sound instructional design;
- taking into consideration the benefits offered by distance teaching (e.g. allowing the student to study in his own home, at his own pace and using media to which he has easy access);
- using multimedia study packages.

Following the principles of instructional design, the teaching programmes were first tested on smaller target groups (which are readily available) to provide feedback. The Library started with in-service training programmes for its own staff, followed by programmes for academic and research staff. Each course is evaluated by the participants. The Department workshops were evaluated and since 1996, students have been asked to evaluate the multimedia study package as a whole. This is done according to certain criteria which should also introduce students to the teaching principles underlying the design of the programme (e.g. the logical presentation of the content, media used, the use of performance activities, the inclusion of activities where the student had to do additional reading and answer some questions, the use of self-evaluation questions to monitor progress, and the support and motivation provided to students).

5. A multimedia study package

A multimedia study package was selected as the most appropriate after analysing:

- the subject content related to online searching and the skills involved;
- the characteristics of distance teaching;
- the needs and requirements of adult students (e.g. for subject related examples, studying independently and having the minimum imposition on their freedom and time).

Two approaches can be followed in defining multimedia study packages, namely:

- multimedia seen as a specifically designed way of combining sound, still and motion images, graphics and animation, and data and text together with the interactive capabilities of a computer (Ref 12);
- multimedia regarded as a teaching approach.

We consider multimedia as a teaching approach which also allows for the incorporation of the first-mentioned idea of multimedia. Looking at a multimedia study package as a teaching approach, the following characteristics should be acknowledged:

- it is an independent self-study package;
- it guides the student's independent activities (i.e. it provides guidelines for the use of the study package, the completion of activities, etc.);
- a variety of media is combined in a meaningful and well balanced manner;
- a number of learning events are included (watching a video, working through a CAI tutorial or attending a workshop);
- the student can monitor his own progress and knowledge of the work by means of self evaluation questions;
- the teaching institution supports the learner (by providing opportunities for mastering computer skills, improving access to technology and answering questions, etc.);
- each medium (a video, study guide or workbook) should be optimally used according to its characteristics (by providing visual displays or demonstrations or practising a certain skill);
- modules are used to allow for different interests and backgrounds.

When designing a multimedia study package it does not matter whether the study material is delivered in printed or in electronic form by means of the Internet. The emphasis is on *how* the media will be combined to support the above-mentioned characteristics. A variety of media are available with guidelines for selecting and combining them (Ref 5). It is important to bear in mind that only technology to which the target group has access can be used.

At Unisa we follow an approach largely based on the printed word. Other media are however also incorporated with the final requirement, namely that the student must attend a short practical session of not more than a week. As the availability of technology increases it will be possible to present workshops all over the country. Once students have access to the CD-ROM databases and the Internet from their homes, it would also be possible for them to complete the practical exercises on their own. It is however still important that they can call on the teaching institution for support.

6. Proposal for a multimedia study package

The proposal is based on sound instructional design which we considered as a non-negotiable prerequisite, considering the characteristics of both distance education and multimedia.

The following core topics apply to all the target groups: the process of information searching; identification and analysis of information needs; process of online searching; online industry and its components; databases and database categories; database structures; search strategy formulation; techniques for online searching; commands and protocol for specific products or forms of online searching (e.g. CD-ROM searching and Internet searching); evaluation of search strategies; database design; use of database programmes.

In addition to these, practising librarians and Information Science students learn about the evaluation of different products (databases, database vendors, database programmes and information retrieval systems); problems related to online searching and how to solve these; information services and the training of end-users. More advanced applications concerning the Internet (e.g. home page design) could also be added. The specific content and the level at which it should be presented will then be decided upon according to the target group, (academic staff, undergraduate students or Information Science students).

The order of presentation for the content should be based on a specific approach or teaching model. For Information Science students the Department uses the database approach as point of departure and with the library users the information need or research process is taken as point of departure.

Every multimedia study package would have an overall goal (i.e. what it wants to achieve), used to formulate the aims and performance objectives. The distance learner should always know what is expected of him and he should be able to monitor his own progress. For online searching, three types of performance objectives can be

distinguished, namely: what students should know (cognitive skills), what students should be able to do (psychomotor skills) and what their attitude should be (affective skills). As has been pointed out earlier on, it is very important that practising librarians and Information Science students should have a positive attitude toward ongoing developments in the online industry and that they should realise the importance of teaching skills. All students should also know the meaning of the concept 'online searching' and the different media used; they should be able to interpret a database structure and they should know how the database structure will influence the search results; they should know how to use one or more command languages or menu-driven systems and they should be able to use these systems on their own. They should also be able to formulate a search strategy. The above-mentioned are all based on theory which the student can master by means of printed media (e.g. study guides or workbooks). This can also be supplemented by the use of videos, video conferences including demonstrations, or computer-assisted tutorials which allow the student to practice certain skills until they are mastered.

As part of the study package there should be opportunities for students to monitor their own progress by means of self-evaluation questions. It is however very important that answers and explanations are provided as well. It should be remembered that the teacher is not physically present to answer the student's questions and the teacher should anticipate any questions that may arise. Learning is also enhanced by the active involvement of students. Tasks can be set and once again the tasks will vary according to the needs of the specific target group. Information Science students are, for example, asked to join the local online user group and to report back on its activities. Other examples of active involvement include the actual evaluation of databases or database programmes, reference interviews aimed at online searches and the evaluation of the study packages according to a set list of criteria. Activities that might be of interest to the other students include the analysis of information needs, formulation and adaptation of search strategies and the design of a data structure to meet their personal information needs. Although it will certainly enhance their understanding of the subject if they could practice the work on a computer, the use of videos and demonstrations can still enable them to master the basic principles — in their own time and in their own homes.

Experience has shown that although there are many methods of supplementing the teaching of online searching, nothing can replace the value of practical exercises on the real product. A workshop or a suitable alternative replacement (e.g. students completing their practical exercises on the CD-ROM network) is therefore still an important part of the training of online searching. The student should however be able to master the basic knowledge at home. He should practise basic skills by means of the set activities and monitor his progress by means of self-evaluation questions, and should also be evaluated by the teacher to ensure that he has mastered the theoretical work before attending the workshop.

The workshop can cover different methods of online searching such as CD-ROM searching, Sabinet searching, Internet searching and database design. The emphasis should be on the practical searches and on evaluating the research results and improving the search strategies. Theoretical knowledge of evaluation is insufficient; it should be based on actual experience with an online session. With a week of full day exposure to practical online searching they should, however, get the necessary experience to master online searching skills. The examples for the practical exercises should be based on the students' subject interest — a point which was strongly emphasised by the academic staff at Unisa.

A selection of media is used in the Unisa study packages, for example study guides and workbooks (e.g. for specific databases or database systems or database programmes such as Inmagic). Videos on online searching are recommended to students and a computer-assisted tutorial is used to teach search strategy formulation. There are also a number of computer-assisted tutorials available from specific database vendors that can be incorporated in the study package.

Methods for evaluating students are included such as assignments and projects. Evaluation before attending the practical session is especially important. The study package as such should also be evaluated. At Unisa, feedback is received from the Information Science students and from Unisa staff who attend workshops.

A final aspect to bear in mind for any distance teaching programme is that it is aimed at adult learners who are often also working full-time. Their needs, interests and the value of their time should be respected. They should also get the opportunity to apply their personal experience, for example by completing subject-related practical exercises and by designing personal databases. This implies that an adult teaching approach will be used.

8. Conclusion

Distance teaching offers valuable opportunities for the education of adult students unable to attend conventional classes. A multimedia study package would include the use of printed material as well as videos, computer-assisted tutorials, workshops and the Internet. The basic theoretical principles of online searching can be mastered before students attend practical sessions. The success of the programme does not lie in the combination of media only but in sound instructional design that will allow the needs of different target groups to be recognised and catered for. The characteristics of distance teaching and multimedia teaching approaches are of special importance. The programme should make allowances for the needs of people from developed as well as underdeveloped environments.

Dr Ina Fourie
Department of Information Science
University of South Africa
PO Box 392
Pretoria 0003
South Africa
Tel: +27 (0)12 429 6313
Fax: +27 (0)12 429 3400
E-mail: fourii@alpha.unisa.ac.za

References

- [1] Barron, D.D. (1987) Alternative delivery of library and information science education: introduction, *Journal of Education for Library and Information Science*, **27**(4), 219-222.
- [2] Barron, D.D. (1987) Faculty and student perceptions of distance education for library and information science, *Journal of Education for Library and Information Science*, **27**(4), 258-271.
- [3] Barron, D.D. (1991) Distance education and the closing of the American library schools, *Library Quarterly*, **61**(3), 273-282.
- [4] Dick, W. and L. Carey (1990) *The Systematic Design of Instruction*, 3rd edition, Harper Collins, Florida.
- [5] Fourie, I. (1994) *Ontwerp van 'n multimedia-studiepakket vir die afstandsonderrig van gerekenariseerde inligtingherwinning*, DLitt el Phil thesis, Rand Afrikaans University, Pretoria.
- [6] Garrison, D.R. (1993) Quality and access in distance education: theoretical considerations, in D. Keegan (Ed.) *Theoretical Principles of Distance Education*, Routledge, London, pp. 9-21.
- [7] Harley, G.S. (1982) *Onderrigontwerpmodelle en onderrigmodelle in tersiêre onderwys met verwysing na afstandsonderrig*, DEd thesis, University of South Africa, Pretoria.
- [8] Haythornthwaite, J. (1990) Distance education and the information scientist, *Aslib Proceedings*, **42**(1), 31-39.
- [9] Holmberg, B. (1991) *Theory and Practice of Distance Education*, Routledge, London.
- [10] Keegan, D. (1983) *Six Distance Education Theorists*, Fern Universität, Hagen.
- [11] Keegan, D. (1990) *Foundations of Distance Education*, Routledge, London.
- [12] Lunin, L.F. (1992) *Multimedia in the Information Industry*, National Federation of Abstracting and Information Services, Philadelphia, PA.
- [13] Peters, O. (1993) Understanding distance education, in D. Keegan (Ed.) *Theoretical Principles of Distance Education*, Routledge, London, pp. 10-18.
- [14] Seels, B.B. (1995) *Instructional Design Fundamentals: A Reconsideration*, Educational Technology Publications, Englewood Cliffs, NJ, p. xi.



Creating a global stock exchange: does the present national system have a future?

Charles Warner-Allen

Perfect Information Ltd, UK

Conjure up a stock exchange in your mind's eye. Typically you envision a neo-classical building in the heart of the financial district with a trading floor crammed with besuited young men frantically shouting and gesticulating among bits of torn paper and surrounded by flickering screens. The images may alter with each city or part of the world but the one constant feature in all of them will undoubtedly be the screens.

Watch those screens carefully. With their high-powered telecommunications links, they are the central nervous system of the world's financial markets which snake their way from Tokyo through London to New York like a main circuit-cable. Without them, the markets won't function. A decade ago the screens may still have been there but they were video screens, not digital. While capable of informing, they were incapable of trading. Trading was effected through personal interaction, comprising handshakes, and seemingly quaint traditions like the London Stock Exchange motto — 'my word is my bond' — were in fact, the essence of deal making.

Others will recall the stock exchange in almost every large British city. Liverpool, Birmingham, Bristol and Glasgow all had exchanges dedicated to meeting the capital needs of the big companies associated with the region's commerce.

It is only this year that settlements of London's stock exchange deals have become fully automated and full electronic trading has yet to arrive. It was only weeks after the 1986 introduction of SEAO — Stock Exchange Automated Quotations — that the first major computer crash rendered dealing impossible. But with the benefits of hindsight, those attempts at market-making using networks of ridiculously under-powered personal computers seems a little naïve by today's standards. And having gone through this much change, surely it would be equally naïve to expect the status quo to last?

Just what is this 'status quo'? Three major stock markets strategically located across the world in New York, London and Tokyo supposedly allow for round-the-clock trading. But this only applies to the stocks of a relatively small group of truly global companies which sport global brands — including Sony, Volkswagen, BAT Industries and Microsoft. While the United States is far and away the biggest of the three markets in terms of market capitalisation, the London Stock Exchange could be described as the most global of the three. In 1995 the market value of all shares in the UK represented 125% of national income, compared with 77% in the US. The figure for Germany is just 24% while Tokyo's share is also small by comparison.

The ever-increasing globalisation of the world economy and the companies within it has been the main factor behind the polarisation of equity trading centres. The late 1980s saw a spectacular increase in mergers and acquisitions activity with hostile takeovers, leveraged buyouts and asset stripping perfected by such practitioners as Hanson and KKR. Recorded UK mergers and acquisitions reached as many as 1528 in 1987 before dropping to below 1000 in 1990. Whole new information services sprang up devoted to recording M&A activity, from the FT to IFR Securities Data and Corporate Money.

The political and economic impact of the changes over the past decade, including the collapse of the Soviet Union and the creation of the single European market, cannot be underestimated. While other regions of the world form trading blocs, such as NAFTA in North America and APEC among nations of the Pacific Rim, they must be set within the context of the economic globalisation in order best to be understood. While the nations of Europe feverishly debate the relative merits of establishing a pan-European federal state, global superpowers like the US seem strangely disinterested. But with the increasingly global nature of the world economy and the fact that the commercial power of many companies supersedes that of many countries — note that a recent study by UNCTAD found that one-third of world trade takes place among parent companies and their foreign affiliates — Americans can be forgiven for looking upon the formation of a regional trading bloc as passé.

In much the same way as Britain's regional exchanges became eclipsed by London it should be expected that emergence of the Big Three would start to mean the decline of other, smaller stock exchanges throughout the world. But in actual fact, national and regional stock exchanges remain in nearly every financial centre, from Berlin to Bratislava and from Chicago to Guandong. The 1995 *Euroclear IFR Handbook of World Stock & Commodity Exchanges* lists some 250 with the number growing. Only last August two Malaysian companies, STI and Petra Khyra Investments, agreed to set up a stock exchange in the Russian republic of Tartarstan.

Telecommunications, in combination with reliable computerised dealing and settlement systems, are the other side to the global stock exchange equation. Screen-based trading — which has reached markets as far away as Bombay and Istanbul — offers the speed, transparency and efficiency required in today's global economy. But not every public company starts out by seeking an international presence. Many simply require

Online Information 96 Proceedings

Page 93

efficient access to capital in local currencies to local finance growth and have no need for the expenses of a global stock exchange.

Finally, the Big Three stock markets all face competition. The principal American stock market, the New York Stock Exchange, is not alone. It has rivals in the much smaller American Stock Exchange and particularly NASDAQ — the National Association of Securities Dealers Automated Quotation system. This deals mainly in smaller over-the-counter stocks and, significantly, has been reporting bids and offers electronically for a quarter of a century. The challenge to Tokyo may be more potential than actual but Singapore is currently in the process of turning itself into a wired community and looking hungrily for a world financial role.

London is facing the biggest challenge to its primacy in the European time zone because of its inability to commit itself to the process of European monetary union. Not surprisingly, both Frankfurt and Paris are claiming that, if it is not prepared to participate, there can be no continuing global role for London as a financial centre with regard to foreign exchange markets.

However, the pluralism and diversity of the world's many stock exchanges is both a strength and a weakness. The last decade has seen the growth in Europe of new secondary stock. London now has the Alternative Investment Market, which aside from the sort of hiccups one would expect when floating newer, higher growth, higher risk companies, has successfully replaced the former Unlisted Securities Market with a more viable source of public equity. Paris has its rival *Nouveau Marché*, which is linked to Frankfurt's *Neuer Markt*. Meanwhile the Belgian, French and German exchanges are establishing, through a series of parallel listings of national exchanges aimed at innovative high-growth companies, something called Euro NM.

But the most formidable of the new European exchanges is the new pan-European EASDAQ, which has taken advantage of the 1995 Investment Securities Directive and the Mutual Recognition Directive issued by the European Union allowing European companies to issue prospectuses in any member state. EASDAQ is aimed at helping fast-growing companies — particularly high-tech and biotech companies — raise capital from a broad pan-European investor base.

EASDAQ has particularly strong systemic credentials. Its automated-quotation trading system will be run by the International Securities Markets Association (ISMA) and will be modelled on its existing trade confirmation system (TRAX) which is used successfully for the Eurobond market. The exchange will be quote-driven by market-makers but includes an in-built ability for periodic order-matching of securities. Clearing and settlement will be provided by Intersect, an established system used by Swiss banks. Finally, the exchange's regulatory framework will be modelled on the US's Securities and Exchange Commission so EASDAQ is drawing off of as much existing experience as it can, which seems sensible.

While these numerous, competing and often conflicting smaller markets have their problems, the net result will be to increase competition amongst the equity markets which will enhance fund-raising options for smaller to medium sized, higher growth companies. As it is sectors such as high-tech and biotech which have not developed as successfully in Europe as in the US, it is hoped that the development of such exchanges will change this.

Local stock markets can facilitate and accelerate growth in emerging markets. In eastern Europe, for example, every economic liberalisation begins with privatisation of state enterprises. With a legacy more often than not of chronic over-staffing and under-investment, plus an almost total lack of market strategy, these companies may not yet be of interest to international investors. So a stock market operating in the local currency — which may not necessarily be a convertible one — is a *sine qua non*. And the fact that both the technology and the regulatory systems are likely to have been imported from the West will tend to create *de facto* world standards which could in due course ease the transition to global trading.

But what the present world set-up fails to do is take full advantage of rapidly falling costs and increasing capacity in both computing and telecommunications. It is plainly not the most efficient way of doing things if you are a multinational company seeking access to footloose capital, wherever it may be held currently, to finance activities that could take place anywhere on the globe. The present limited international arrangements more resemble a relay race, with the baton passed from one market to another as the sun travels round. And, just occasionally, the baton gets dropped.

Take Black Monday — 19 October 1987. Less than a year after London's Big Bang, a sharp fall in share prices in New York was exacerbated as first London and then Tokyo reacted by opening lower and closing even lower still. Could it have been avoided if all trading during that fateful 24 hours had taken place, not in three stock markets but in one? Possibly not — stock exchanges are febrile places and, at the time, at least part of the blame for the collapse was put on programme trading, which could presumably flourish in a single global exchange. But the fact remains that what we have now is not the most efficient arrangement, bearing in mind what technology could offer.

So if we already have something approaching a global stock exchange — albeit a tripartite one — and the technology available to permit total integration at a reasonable cost, why isn't it happening? There are three main areas which need to be investigated.

Technological issues

You only need to think of TAURUS to realise that we have by no means cracked all the technological problems. Having established an effective system for reporting trades, the London Stock Exchange then looked at computerising the final settlement of deals and came up with TAURUS — Transfer and Automated Registration of

Online Information 96 Proceedings

Page 94

Uncertified Stock. The basic idea was to reduce the need for handling share certificates by letting a computer system act as a clearing house. But it was an unqualified disaster which was abandoned in 1993, to be replaced in July of this year by Crest — not a moment too soon, because all of London's international competitors have paperless settlement systems.

Even so, it's going to take until April next year to transfer all securities from the present manual Talisman system. At the moment, armies of clerks play 'Snap' with cheques from buyers and receipts from sellers. To the outsider it must seem extraordinary that with a decade's experience behind us in high speed computerised trading in so many different kinds of financial market, it should have proved so difficult to come up with a viable settlement system. But it did — and there's a moral here for the creators of a global stock exchange.

Also newly arrived, at the end of August, was the London Stock Exchange's Sequence trading system. A decade after the Big Bang, dealers are still reporting details of trades by phone instead of entering them into the computer themselves, and full electronic trading is still not due to be implemented until next year.

There are more lessons to be learned from the general state of global telecommunications at their present stage of development. Not for one moment am I suggesting that market makers in a global stock exchange would use the public version of the Internet. That would be akin to making the police catch the bus every time they wanted to chase someone. Nevertheless, the chronic congestion and at times tenuous links on what might eventually become the information superhighway should carry warnings for the creators of the global area network that would be necessary for a world stock exchange. Here in Britain we're currently facing the prospect of the third wholesale reorganisation of telephone numbers in less than a decade.

The technological infrastructure is there: it is a question of how much credibility it can garner from investors and companies. Until all the participants have the confidence that the system is secure against abuses, and that it is capable of handling the immense bulk of securities required of it, the technology will remain limited by the people behind it. Common sense suggests that it would be foolish to create a system that didn't use world standards. At least part of the operations of a global stock exchange would need to be potentially accessible to millions of people who might only use it occasionally. I'm thinking in particular about the global equivalents of TOPIC or the Regulatory News Service, where market transparency demands that prices and company announcements be disseminated instantly as widely as possible. The same rules would have to apply, too, to the images of documents such as Perfect Information distribute now. The moral is that however big the network is, it won't be big enough.

In September 1995, two California University students broke a new Internet encryption system almost as soon as it was launched by working out how to predict the random numbers that Netscape allocated to online credit card transactions. Internet front runners such as Netscape and Microsoft are talking about effective encryption technology by next year. But first of all you have to allow for the tendency in the industry to talk up developments that subsequently don't quite deliver what they promised. And, in any case, those base level standards won't be good enough for a global stock exchange. To gain any credibility it will have to be demonstrably secure, while at the same time offering worldwide access to its services using world standard protocols. That's a tall order.

The same exacting standards will have to apply to the settlement system. Again, the pundits assure us that secure payment systems via the Internet are a matter of a few months away, and maybe they are. But let's not fool ourselves that there won't be teething problems. Again, the bulk standard security arrangements for the public Internet will not be good enough for a global stock exchange. It will need to be like Caesar's wife — above suspicion. Most current corporate fraud never comes out into the open because of the potential damage that might be done to the victim company's credibility. Can you imagine the damage that would be done if the same were true of a global stock exchange? For every fraud that did come to light there would be suspicions of a dozen more, and that could be fatal. Again, therefore, the settlement system must meet the twin requirements of transparency allied to security.

But let's be positive and say that the technological problems are short term ones. There's no shortage of straws in the wind. Entrepreneur Peter Levin launched in the spring a service called Display.IT, which uses the Net to distribute financial data at a claimed cost saving of 90% over conventional feeds such as Reuters. Ofex, the tiny unregulated London market, posts its mid prices free on the Net. About now, the Deutsche Borse is to start offering Internet banking services, allowing customers to query securities accounts and place orders for securities transactions.

Regulatory and institutional issues

But hand in hand with the technology go regulatory and institutional issues. There are currently scores of independent stock exchanges serving radically different markets. And if we believe that our current systems of financial regulations are the best that could be devised, we should keep an open mind. We've seen two recent examples of breakdown — the Nick Leeson affair, in which a Singapore based trader brought down the venerable Barings Bank and, more recently, revelations of losses at Deutsche Morgan Grenfell's unit trust business as a result of 'bizarre' investments by Peter Young.

There's been a measure of harmonisation of securities legislation in the European Union over the years, with Directives covering issues such as minimum standards, mutual recognition of competent authorities, and a European 'passport' for securities dealing activities. Measures such as the Admissions Directive, the Listing Particulars Directive and the Interim Reports Directive date back to the late 1970s and early 1980s, and progress generally has been slow. Stock market regulation is fundamentally a national matter, and rules and regulations

vary widely from country to country. For a global stock exchange to be perceived as being effectively regulated, it would have to meet the requirements of perhaps dozens of different legislatures. Individual governments would be perfectly capable of imposing onerous additional requirements on local dealers if they were not satisfied with the regulatory arrangements, to the detriment of the global exchange's operation.

The establishment of EASDAQ offers one example: Many practitioners expressed concern over differences regarding investors' rights in different countries. European investors used to UK-styled pre-emption rights with their focus on shareholder value are worried that companies listed on EASDAQ will be able to waive such rights by convening an EGM whenever matters deemed significant to the company arise. Equally, US companies with a dual listing on NASDAQ and EASDAQ — which EASDAQ is keen to attract — will be frustrated with the need to convene an EGM in the first place because they are unused to having to deal with pre-emption rights at all. Thus, a cross-border exchange will face challenges in creating a single set of rules that appeals to investors from every country.

Looking at the issue of European monetary union, other problems have become evident. France and Germany have been claiming that London banks should have to pay a premium to carry out transactions involving the new currency. While the main impact would be on London's foreign exchange markets, such restrictions might also compromise London's position as Europe's financial centre. Lawyers have also been discussing the possibility that the arrival of the single currency could annul certain cross-border contracts, suggesting that swaps and derivatives contracts could be challenged under UK law if both the currencies involved disappeared. The Bank of England's deputy governor, Howard Davies, denied last September that the City's premier financial role would be damaged by membership or non-membership. Nevertheless the EMU issue currently seems to be splitting industrial and financial sector leaders in Britain, and the lesson for a global stock exchange is that both national and supranational law, over which it would have no direct influence, could affect its operations.

If you look at the current Big Three stock markets, even their regulatory arrangements differ in important particulars. London has a competing market maker system with full electronic display of bid and offer prices. Tokyo trades its most active stocks on the stock exchange floor and it has regular member firms, who act as brokers and dealers, and so-called *saitori* members who function solely as intermediaries in transactions between regular members. New York has a wide diversity of participants including member firms who can function as both agents and dealers, and also as brokers on the trading floor, executing customer orders; also participating in New York are independent brokers, specialist dealers in individual securities, and institutional and individual investors. Round the clock trading only works because these three quite different markets are more or less in the right place to facilitate global trading, because they are big, and because they are internationally perceived as being reasonably safely regulated.

NASDAQ is the trading system for the United States over-the-counter market, dealing largely — but not exclusively — in smaller company securities. It does also trade in large company stocks, including those listed on other stock exchanges. Of its 180,000 terminals in use in 1995, about 25,000 were outside the United States. NASDAQ was designated a United Kingdom recognised investment exchange in 1988, under the Financial Services Act 1986; NASDAQ International has been operating since early 1992 and it allows investors to trade United States stocks before the New York Stock Exchange opens.

NASDAQ also has a sister exchange in EASDAQ — which we have already discussed — in which it has a 10% stake. The two exchanges, although separate, will use very similar trading, settlement and regulatory systems and will encourage companies from North America and Europe to seek a dual-listing to take advantage of investment opportunities on both continents.

A further possible model is the Tradepoint Investment Exchange, a tiny rival to the London Stock Exchange. Last February it announced a deal with the data and information provider Bloomberg, whereby users of the 55,000 Bloomberg terminals worldwide would be able to use them to execute trades in UK shares. Tradepoint's executive director Stephen Wilson acknowledged at the time that this would give his exchange a global reach that it could not have built for itself. And these are early days for Tradepoint anyway. It lost £5.7 million in the year to March 31, gaining only £77,706 in trading income in its first eleven months.

Nevertheless, the lesson is clear — given business confidence and political will, international stock markets can be created.

Sovereignty and national pride

Much more problematic than the regulatory arrangements, I suspect, are the more nebulous issues of sovereignty and national pride. As we've already seen, a stock exchange is one of the first institutions that a newly emerging market economy wants to create — and not merely for economic reasons. The relatively small number of quoted companies tend to contribute disproportionately to their country's gross domestic product. Consequently, stock markets are perceived politically as being a vital weapon against competitor countries.

As many continental European companies face up to the need to increase their proportion of equity funding to foster growth and compete in a global economy, they will also have to face up to the prospect of foreign ownership of their companies. This has proved to be a particularly galling prospect for French companies which are said to be energetically supporting the development of equity markets on the continent — witness the many we have so far discussed — as a means of avoiding what they perceive to be the 'humiliation' of Anglo-Saxon dominance in the form of US and UK investors.

Online Information 96 Proceedings

Page 96

The issue of nationality and sovereignty is also central to the debate over European monetary union. Cynics could argue that pro-EMU governments aim to take advantage of the new *status quo* it will establish while those on the anti-EMU side are accused of being regressive and insular by appealing to petty nationalism. Michael Cassidy, chairman of the Corporation of London Policy & Resources Committee, is quite clear about the perceived threat to London that an enlarged Europe presents: 'It is politically driven by the French and the Germans. They are envious of the market share London has developed in recent years.'

So sovereignty and national pride throw up potentially the biggest obstacles in the path of a global stock exchange, being all the more insurmountable because of their basis in emotional politics as opposed to economic logic. Business could in theory create a global stock exchange if it chose. Indeed, as we have seen, there are already models. But a truly global exchange — a single institution, operating electronically in every world region — would still have to be subject to the national law of the countries where its dealing terminals were located, and some politicians could be obstructive.

How could a global stock exchange come about? Firstly, the Big Three — in New York, London and Tokyo — could consolidate their current baton-passing activity into a more formal, integrated arrangement — possibly creating a new super exchange on top of their existing domestic operations. Naturally such an exchange would have to adopt a single computer system for displaying prices, striking deals and back office settlement. But there would also need to be a single regulatory regime, with one set of trading rules dominating over those of individual nations. As we have already seen, current arrangements in the three countries are quite different and the creation of a single regulatory regime would require the acquiescence of governments in three competing economic blocs to the domination of a single system representing the most successful option. Deciding on and establishing such a system presents many tricky problems.

One of the Big Three could come to dominate globally. The US market capitalisation is twice that of Japan and five times that of the United Kingdom and should dominate on that basis. But in terms of foreign equity trading, London has well over half the market share, New York another third and the rest of the world picks up the balance. This gives London a genuine international edge over New York, which is enhanced by its time zone positioning in that it is possible for traders from London feasibly to trade with those in the Far East and the US, being positioned between the two.

Thirdly, an existing smaller exchange could start to compete aggressively in the international market, taking advantage of technological developments, its existing domestic expertise and a supportive regulatory environment. But quite apart from having to win international credibility, any likely candidate would be coming from a long way behind. Freewheeling Hong Kong may be sixth in market capitalisation but at a mere 7% that of the United States. Wired and squeaky clean Singapore is seventeenth, at 2.5% that of the States. And if you look at growth rates, in terms of numbers of listed companies, Hong Kong and Singapore are fourteenth and nineteenth respectively. The markets with the greatest increases in listings — Portugal, Indonesia and Turkey are the top three — don't frankly look like front runners for a global stock exchange.

Finally, then, businesses could create a completely new institution, along the lines of the screen-based international currency dealing operations of Reuters or Dow Jones Telerate. But, in regulation and disclosure terms, it's a very long way from dealing in currency to trading the stocks of companies. As the recent hurdles faced by the brave AIM and Ofex show, it costs a lot to maintain the due diligence that is necessary to run a market in company shares. Maybe we have a sort of model in NASDAQ. But NASDAQ is a vehicle for trading US stocks, and its overseas operations require legal recognition in the individual countries concerned. EASDAQ is perhaps the boldest initiative so far, based as it is firmly within the equally ambitious European Union. There can be little doubt that the City of London has been watching its first few steps with great interest. The time may not be right for the establishment of a pan-European or global market yet but we will learn plenty from this pilot operation which will establish the form such an international exchange will take in the future.

So will it ever happen? Technologically, it could happen now; the networks and computing power are available and there is already screen-based global trading in other markets — notably currency. In a sense therefore the rash of interest over the last few years in the information superhighway is not an issue. Global stock exchange trading was feasible before that anyway.

Institutionally, it almost does happen now, through the activities of the Big Three — although this is not a single international operation. Politically, the trend towards the necessarily relinquishment of national sovereignty that would be a prerequisite for a supra-national market is detectable — in the halting progress towards European monetary union, for example, and in the ever-increasing globalisation of business and companies.

All three of these prerequisites — the technology, the institution and the political will — would need to be firmly in place before a global stock exchange actually came about, and that could take some time. All the same, if there's any such thing as a safe prediction, one is that change usually comes faster than expected — whether it be the take-up of the Internet or the demise of the stock exchange trading floor. So let us not be timid in our predictions — a global stock exchange is a question of 'when?', not 'whether?'

And if we bid farewell to the concrete tower of Old Broad Street and the neo-classical facade of Wall Street, how long before we wave their occupants bye-bye as well? Program trading is a crude device at present — the fundamental decisions for its operation are made by people. Blindly applied, it played its part in the crash of Black Monday. But the predictive models that drive it can only improve. Even if computers never beat flesh-and-blood traders in our global stock exchange, they only need to match them to win. But computers will always be reliant on their human creators so the integrity of a global stock exchange system will be determined by the individuals who operate them.

Economic man makes economic decisions based on perfect information. Therefore, the establishment of an

effective stock exchange will only work if it has access to Perfect Information.

Charles Warner-Allen
Perfect Information
35 Chisell Street
London EC1Y 4SE
UK



Advocating global forest issues on the Internet

Alois Kempf

Swiss Federal Institute of Forest, Snow and Landscape Research (WSL), Switzerland

Abstract: Sustainable development, biological diversity and conservation of tropical forests are only a few of the hot environmental and political topics where the actors involved have started to make use of the world-wide computer networks. The Internet (as a transport medium of information exchange) and the World Wide Web (as the favourite service to interconnect related movements) are being chosen by international organisations, scientific institutes, forest products industry and environmental activists for advocating their specific view on forestry, wood production and ecological concerns. The paper points out a few patterns of interaction by presenting Internet services of different institutions involved, for example the Food and Agriculture Organisation of the UN (FAO), the Intergovernmental Panel on Forests (IPF), the International Union of Forest Research Organisations (IUFRO) and the World Conservation Monitoring Centre (WCMC). Special attention is given to information services based on computer-mediated communication, such as the electronic clearing-house Linkages, and to action groups advocating online such as the Rainforest Action Network (RAN) or Ecological Enterprises. The difficult task of information professionals in this context will be to keep track of relevant documents and special interest networks (SINs) on the Internet. Otherwise they will not be able to serve the user in his or her needs for adequate interpretation of accessible documentation from electronic resources.

Keywords: forestry information, Internet, resource discovery, virtual communities, networked information sources, international organisations, conservation activists, information filtering, restricted access, search engines, information quality, user guidance

1. Introduction

'As the end of the 20th century approaches, there is an increasing awareness of the role of forests and trees in human welfare, in our economic development and in sustaining our environment. Forestry issues have captured the world's attention as never before. Phrases like sustainable forest management, deforestation, climate change, biological diversity appear frequently (...). Canada and Canadian foresters have long been leaders in international forestry and have made important contributions throughout the world. More and more, we are looking beyond our borders to understand the issues. The Global Forest, like the Global Village, impacts on all of us' (Ralph W. Roberts at <http://www.infoserve.net:80/forest/global.html>).

This quotation from the International Forestry Home Page produced by the Canadian Chapter of the Commonwealth Forestry Association (CFA) sets the stage well for the increasing interactions observed on the Internet with respect to global forest issues (Ref 1). When I had the idea of preparing this paper it was rather an assumption, then confirmed knowledge, that international forestry was creating a new platform for information exchange through Internet network resources. The evidence found during a few additional visits on the World Wide Web in May 1996 made it clear that computer-mediated communication and the presentation of various interest groups are not of accidental. Glen Barry for example explicitly states in 'Using informational technologies for forest advocacy' (29 January 1996) that 'conservationists have no options but to embrace informational technologies at this critical junction' (see Ref 2). Information professionals and end-users may welcome this media shift of pressure group activities towards a globally accessible electronic forum because dissemination of news, facts, visions and events are fast and theoretically ubiquitous. The common transport channel makes it possible to build new partnerships and share knowledge within multinational working parties. But what is the impact of the virtual 'Global Village' on the 'Global Forest' on earth, to paraphrase the CAF statement above?

2. Methodical approach

As well as its public functions, the Internet can serve as a toolbox for invisible interaction between individuals or within closed groups, as documented for the Consultative Group of International Agricultural Research CGIAR (Ref 3). Thus, information resources that are publicly available via the Internet are without doubt only the tip of the iceberg. This is important for the following description of forest advocacy cases which are based on openly accessible documents and services. The term 'advocacy' is used in its large sense to cover informational activities on a specific subject, i.e. global forestry. Due to the dynamic of the Internet services and the limitations of available network tools for the analysis of on-going changes and the mapping of the current interactions in a subject field, the chosen examples are only pointers for further retrieval and personal observations. To locate networked sources on advocacy for forest issues is essentially asking 'Who does what, where, when and why in reality?'. Virtual libraries (Ref 4), subject-oriented metadatabases (Ref 5) and lists of links on relevant hosts (Ref 6) have served as starting points for a focused network navigation in order to learn more about the actors in the field (Refs 7, 8, 9). Using the advanced features of existing search engines — preferably Alta Vista in our case — helped a lot in further following up traces of organisations, events or abbreviations discovered in the context of global forestry.

3. Facts and opinions on global forestry

'Geneva, Switzerland — While governments have spent the last two weeks talking inconclusively about forests, more than half a million hectares of forest have been destroyed worldwide' (WWF News release, 22 March 1996, about 2nd IPF meeting).

A lot of forest-related directory and bibliographic information concerning associations, institutions, experts or projects is offered via Internet services. Full-text versions of conventions, agreements of relevance to international forestry as well as reports about the state of world forests can nowadays be looked up online. Table 1 shows a small selection of the various document types found on the Internet (May/June 1996).

Table 1: Examples of networked information sources on global forestry.

Products/services	Internet reference (uniform resource locator)
Bibliographies at the University of Minnesota Forestry Library	gopher://minerva.forestry.umn.edu/11/
Oxford Forestry Institute, <i>Tropical Forestry Papers</i>	http://ifs.plants.ox.ac.uk/OFI/pubs/tfppubs.html
IUFRO Newsletter	http://iufro.boku.ac.at/news/newslet.html
Forest Ecology and Management, Elsevier's Tables of Contents Service	http://www.elsevier.nl/cas/estoc/contents/SAH/03781127/SZ955019.html
Directory of Forest Pathologists and Entomologists	http://www.ncfes.umn.edu/pubs/pdf/worlddir.pdf
US Forest Service, Phone Directory	http://www.fs.fed.us/intro/directory/welcome.htm
Current projects, compiled by ETRN	http://www.dainet.de/etfrn/pro.htm
Forest Conferences, from ERIN Australia	http://www.erin.gov.au/land/forests/conference.html
TROPIS Data Entry Form, from CIFOR	http://www.cgiar.org:80/cifor/tropis/tropisform.html
Agenda 21 (Rio'92), Chapter 11 on Deforestation	gopher://gopher.undp.org:70/00/unconfs/UNCED/English/a21_11
World Rainforest Report, from Rainforest Information Centre (RIC), quarterly	http://www.igc.apc.org/ran/info_center/wrr/index.html
World Resources 1996-97, A Guide to the Global Environment; Chapter 9 on Forest and Land Cover	http://www.wri.org/wri/wr-96-97/index.html
FAO State of the World's Forests	gopher://faov02.FAO.ORG:70/11Gopher_root%3A%5Bfaoworldfo.T1FF%5D
FAOSTAT, a facility to produce wood statistics on demand from a database	http://www.fao.org/waicent/forestry.htm
Countdown Forests 97, International Forest Policy, Issue 1/March 1996	http://iisd1.iisd.ca/forests/countdown/c97-1.htm

Online Information 96 Proceedings

In harmony with the objectives of an organisation there must be a bias in the selection, compilation, dissemination and presentation of specific forest-related issues. Thus, pluralism in data interpretation and knowledge proliferation is expected. Consequently it is up to the user to decide what he or she thinks is right in a particular context. In this respect the situation does not differ from the one of a visitor in a traditional library or documentation centre. But improved access through electronic publication on the Web or due to accelerated document delivery to original documents which are otherwise barely available has contributed in forestry to intensive discussions and to an increase of participants involved. Explicit statements made by the different parties about their organisation, mandate and vision are important for outlining a sociogramme of computer-mediated interaction. To continue the above example in Barry's article, the case of the Ecological Enterprises Mobile Awareness Team in Papua New Guinea (PNG) follows (Ref 2):

'The objectives of Ecological Enterprises' awareness patrols are:

1. To help facilitate the protection of still intact tropical ecosystems and cultures of Papua New Guinea.
2. To provide information on environmental effects of industrial logging to the legally recognised owners of the land.
3. To encourage debate and discussion over what type of development options are best for the long term good of the Papua New Guinean people.
4. To provide counsel to the landowners on the issues of legal rights, making contractual obligations understood through advocacy work and explaining the procedures for logging negotiations.
5. To empower and support lasting local environmental initiatives. It is critical that indigenous people are able to gain skills and information to build their own grassroots movement.'

Similar statements can be retrieved from various decision making committees, task forces and pressure groups in the concert of global forestry (see Appendix). However, the user should be aware of encountering 'virtual personalities' and alias names in discussion groups as well. A summary of the FAO electronic conference 'Addressing Natural Resource Conflicts Through Community Forestry', running from 15 January to 28 April 1996, reports the creation of 'Laurie Amekufa' whose role was to provoke more participation (Ref 10).

4. Virtual interaction and missing links

'... I believe that too often the forestry problem is formulated to fit the available solution, and that forest policies do not make use of the past experience' (From the FOREST mailing list at listserv.funet.fi, June 1996).

Already a 'quick and dirty search' on the Internet makes it clear what the hot topics in forest advocacy presently are: biodiversity, boreal forests, certification and labelling of wood, deforestation, old growth, rain forests, sustainable forest management, tropical forests: in short, the recall covers mainly areas within the more general controversy on ecology versus economy. Therefore, information sources found on these issues also originated beyond the traditional forestry sector which, on the other hand, had integrated many of the environmental concerns in its old silvicultural concept of sustainability for decades. While some of the items, such as old growth, are of greater interest in North American and European regions, other topics are heavily reflecting aspects of the North-South development conflict. The history of CGNET (Ref 11), the creation of IUFRONet (Ref 12) and the growth of service providers like the Association for Progressive Communications APC (Ref 13) or the World Telecommunication Development Report 1995 of the International Telecommunication Union (Ref 14) give evidence of a geographically wider use of telematic applications during the last few years.

As a consequence of the technical infrastructure at hand, a wide range of information products were created to support 'networking' activities in the sense of advocating specific subjects: research institutes have contributed publications and project news, forestry libraries are offering bibliographies and document supply services, governmental agencies have started Questions & Answers hotlines, special clearinghouses provide value-added reference centres (Ref 15). Linkages, from the International Institute for Sustainable Development (Ref 16), runs among other services the Global Forest Policy Clearinghouse which stores a lot of important documents about the Helsinki Process (since 1990) and the Montreal Process (starting 1993), and information on the UN Intergovernmental Panel on Forests (IPF) or on the Forest Stewardship Council. The European Topic Centre (ETC) for Forest, a focal point mechanism of the European Union, is still in the planning stage (Ref 17).

'Alert services' and action oriented 'Archives' are two favourite network resources of information which is for public downloading or dissemination by many non-governmental organisations or regional activist groups. Table 2 lists a few of them selected from different forestry information meta-indices (Ref 6). Further data collections and forest-related movements can be found via the uniform resource locators (URL) given. The IRIS Forest Products Industry Directory, produced by Forest Products Industry Consultants and the World Timber Network and managed by Lejé International (Ref 18), contains many links referring to the forest industry sector (Ref 19).

Table 2: Internet resources of some environmental advocacy actors.

American Forests, Global ReLeaf	http://www.amfor.org/programs/plant.html
EnviroLink Library, Forestry Issue	http://www.envirolink.org/EnviroLink_Library/Flora_and_Fauna/Plants/Plant_Wildlife/Forestry_Issues/
EnviroLink, Express Yourself	http://www.envirolink.org/express/
GAIA, Gaia Forest Conservation Archives	http://forests.lic.wisc.edu/gaiaarc.html
GAIA, Worldwide Forest/Biodiversity Campaign News Archives	http://forests.lic.wisc.edu/worldfor.html
Greenpeace International, Interactive Chat	http://www.greenpeace.org/icha.html
Rainforest Action Network, Action Alert Index	http://www.igc.apc.org/ran/info_center/aa/index.html
Sierra Club, Action Alerts	http://www.sierraclub.org/sc-action/
World Wide Fund for Nature, Global Action Alerts	http://www.panda.org/action/actmain.htm

Very popular features in the context of forest awareness promotion are 'mailto:' buttons, built-in for various purposes, and WWW forms as pre-produced reply documents. Discussion lists, bulletin boards or newsgroups are other frequent patterns of information exchange. As far as global forestry is concerned one can observe that different organisations have started electronic conferencing on similar topics. One example is the Global Web Forum of the Malaysian Timber Council (Ref 20) which demands registration to the different lists. Another example, the FAO Forestry Research Education and Extension (FREE) discussion groups (Ref 21), have chosen to moderate and structure electronic interaction through agenda suggestions.

To summarise some of the 'advocacy features' via computer networks, G. Barry (Ref 2), is cited once more. He describes the informational campaign strategies developed by Ecological Enterprises for Papua New Guinea as follows:

- '1. to organise local forest conservation information in a systematic manner to serve the needs of community development and empowerment work in PNG in particular, and in a less detailed manner, the world at large. This involves the collection, selection, compilation, sorting and dissemination of information;
2. to act as support and contact centre for community actions and campaigns;
3. to act as intermediary for receiving and disseminating information;
4. to provide consultation and services to the community and community-based organisations in PNG in particular, and the world in general;
5. to demonstrate to other environmental campaigners how information technologies can be used to network ecological information.'

5. Conclusions

'Should you or your organisation be added to this directory?' (The Sustainable Forests Directory).

The examples above have certainly not shown all the concepts for resource discovery and topic-oriented Internet navigation. Consequently, starting from a geographical approach (e.g. Amazonia, boreal forests, Indonesia) and searching for a specific theme (e.g. 'Santiago Declaration', 'Forest Principles') or for additional networks of special interest groups (SIN, e.g. scientists, activists, governmental committees, local population) would result in further background information. But do we get the right answer to our original question or are we only invited to visit other interesting sites that will shape and disturb our view? Do we in the end know who Glen Barry (Ref 22) is and what the effects of Ecological Enterprises (Ref 23) are on forest conservation in Papua New Guinea?

In my view, some paradoxical phenomena have to be reported with respect to information quality assessment (Ref 24):

- while global forestry has good coverage on the Internet, network access is still very limited from places where forests cover the earth;
- while we are able to locate who is already active on the Internet with respect to forests, there are only restricted tools available to measure the impact of network presence, in order to analyse the inter-linkages of services and to evaluate single documents for their contents' quality level;

Online Information 96 Proceedings

- while networked information resources can heavily contribute to a better definition of the problem, the virtual interaction alone does not automatically lead to problem solving in the field, i.e. in the local forest.

In conclusion, a great challenge of Internet information retrieval on international forestry today is to manage the output of all the major actors from the different regions in the world. As we know, public awareness is still shifting from 'forestry' as an economic sector towards 'forests issues' as a multidisciplinary focus on human land use and forest habitats. Therefore end-users and information professionals will profit from any improvement of searching engines. On the other hand, adding value to information resources publicly available on the Internet will become an important task required from information specialists in the future. To do this properly, new tools for citation analysis of networked information and for impact studies of online interaction are needed. However, one should not neglect that the publicly available 'knowledge base' is only the visible part of the iceberg and does not tell us the complete story of the Internet-based information market.

Alois Kempf

Swiss Federal Institute of Forest, Snow and Landscape Research (WSL)

CH-8903 Birmensdorf

Switzerland

References

- [1] The cited URLs were valid in May or June 1996; there is no way to guarantee their existence as interactive electronic resource during the International Online Meeting in December 1996. See also: Roper, J. (1996) Forestry and the Internet in Canada, *Forestry Chronicle*, **72**(2), 122-123.
- [2] Barry, G. (1996) Building forest advocacy through communication networks in Papua New Guinea. Ecological Enterprises, 22 March 1996. Electronic document at URL <http://forests.lc.wisc.edu/forests/useinfo.html>
- [3] CGIAR Discussion Lists (Open, Announcement, Internal) at <http://www.cgiar.org:80/ivdn/CGLISTS.HTM>
- [4] WWW Virtual Library Forestry, compiled by Jarmo Saarikko, Finnish Forest Research Institute, Helsinki at <http://www.metla.fi/info/vlib/Forestry.html> and the other VL collections in this document
- [5] Benz, J. and K. Voigt (1995) Indexing file system for the set-up of metadatabases in environmental sciences on the Internet, *Proceedings of the 19th International Online Information Meeting*, 5-7 December 1995, London, Learned Information, Oxford, pp. 455-465.
- [6] The Sustainable Forests Directory at <http://www.together.net/~wow/Index.htm>; Environmental Organisations Web directory: Forestry at <http://www.webdirectory.com/Science/Agriculture/Forestry/>; Natural Resources Institutions and Organisations Worldwide, by Yu-Fai Leung at <http://sfbox.vt.edu:10021/Y/yfleung/forestry.html>; CIFOR — Other Links at http://www.cgiar.org:80/CIFOR/info/relevant_links.html; IUFRO — Related Links at <http://iufro.boku.ac.at/>; ERIN (Environment Resource Information Network, Australia), Forest Information Sources at http://www.erin.gov.au/other_servers/category/Forest.html; Canadian Forests — Information and Forestry Links at <http://www.infoserve.net:80/forest/inform.html>.
- [7] Burk, T.E. (1995) Forestry on the Internet: an introduction, *Journal of Forestry*, **93**(5), 6-7, 50-54. Also at <http://smokey.forestry.umn.edu/nres5220/tom/JoF/main.html>.
- [8] Green, C.C. and D.A. Sommer (1995) The World Wide Web: what's in it for foresters, *Journal of Forestry*, **93**(12), 24-29.
- [9] Kempf, A. and V. Brown (Eds.) (1995) Forestry information and knowledge exchange, *Proceedings of IUFRO Subject Group S6.03-00 Information Systems and Terminology*, held at the XX IUFRO World Congress, 6-12 August 1995, Tampere, Finland, Swiss Federal Institute of Forest, Snow and Landscape Research, Birmensdorf.
- [10] Backgrounder on FAO's Electronic Conference on Conflict Resolution in Community Forestry, at <http://www.fao.org/news/1996/960606-4.htm>
- [11] Lindsey G., K. Novak, S. Ozgediz and C. Balson (1994) *The CGNET Story: A Case Study of International Computer Networking*, IDRC.
- [12] <http://iufro.boku.ac.at/iufronet/t1/tfdocs.htm>
- [13] <http://www.apc.org/index.html>
- [14] <http://www.itu.ch/WTDR95/>
- [15] ETRN at <http://www.dainet.de/etfrn/etfrnhom.htm>; Australian Forestry from PIENet at <http://www.dpie.gov.au/dpie/forestryoz.html>; World Forest Institute at <http://www.vpm.com/wfi/>; INFOSouth Resource Centre at <http://wwwfs.libs.uga.edu/>.
- [16] Linkages multi-media resource at <http://www.iisd.ca/linkage/>
- [17] <http://www.eea.dk/eionet/etc.html>
- [18] IRIS Forest Products Industry Directory at <http://www.cdc.net/~primus/IRIS/irisfp.html>; World Timber Network at <http://www.transport.com/~leje/index.html>

- [19] Vlosky, R.P. and R. Gazo (1996) The Internet and the forest products community: the role of the Forest Products Society, *Forest Products Journal*, **46**(5), 19-25.
- [20] <http://www.mtc.com.my/forum/forumacl/dispatch.cgi>
- [21] <http://www.fao.org/waicent/faoinfo/forestry/tree/free.htm>
- [22] Contact address at <http://forests.lic.wisc.edu/forests/current.html>
- [23] ECOLOGICAL ENTERPRISES' PNG Rainforest Campaign, An Informational Support Program for the Rain forests, Biodiversity and Peoples of Papua New Guinea. September 12, 1995. Electronic document at <http://forests.lic.wisc.edu/forests/whatis.html>.
- [24] Clausen, H. (1995) Information quality as a factor in research library Internet use, *Proceedings of the 19th International Online Information Meeting*, 5-7 December 1995, London, Learned Information, Oxford, pp. 61-69.

Appendix: abbreviations and list of organisations (actors) cited.

Abbreviation	Organisation	WWW home page
CABI	CAB International	http://www.cabi.org/
CATIE	Centro Agronómico Tropical de Investigación y Enseñanza	http://www.catie.ac.cr/english.html
CFA	Commonwealth Forestry Association	http://www.infoserve.net:80/forest/cfa.html
CGIAR	Consultative Group for International Agricultural Research	http://www.cgiar.org/
CIESIN	Consortium for International Earth Science Information Network	http://www.ciesin.org/
CIFOR	Centre for International Forestry Research	http://www.cgiar.org:80/cifor/
DPiE	Australian Department of Primary Industries and Energy	http://www.dpie.gov.au/dpie/home.html
ECNC	European Centre for Nature Conservation	http://www.ecnc.nl/
EFI	European Forest Institute	http://www.efi.joensuu.fi/
EIONET	European Environmental Information and Observation NETWORK	http://www.eea.dk:8080/
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária	http://www.embrapa.br/
EnviroLink	Online environmental information resource network, USA	http://www.envirolinks.org/
ERIN	Environmental Resources Information Network, Australia	http://www.erin.gov.au/
ETC	European Topic Centre	http://www.eea.dk:8080/frames/etc/
ETFRN	European Tropical Forest Research Network	http://www.dainet.de/etfrn/etfrn.htm
FAO	UN Food and Agriculture Organisation	http://www.fao.org/
FS-INFO	US Forest Service, Information Network	http://www.fs.fed.us/infonet/
FSC	Forest Stewardship Council	http://antequera.antequera.com/FSC/
FTPP	Forest, Trees and People Programme (FAO)	http://www.fao.org/waicent/faoinfo/forestry/ftpp/default.htm
GEF	Global Environment Facility, World Bank	http://www.worldbank.org/html/gef/Welcome.html
IDRC	International Development Research Centre, Canada	http://www.idrc.ca/
IIASA	International Institute for Applied Systems Analysis	http://www.iiasa.ac.at/
IISD	International Institute for Sustainable Development	http://iisd1.iisd.ca/
IPF	Intergovernmental Panel on Forests	http://www.mbnnet.mb.ca:80/linkages/forestry/ipf.html
IRIS	Industry Related Information Service	http://www.cdc.net/~primus/IRIS/irisfp.html
ITTO	International Tropical Timber Organisation	E-mail: info@itto.or.jp
IUCN	The World Conservation Union	http://www.iucn.org/
IUFRO	International Union of Forest Research Organisations	http://iufro.boku.ac.at/
MTC	Malaysian Timber Council	http://www.mtc.com.my/
OFI	Oxford Forestry Institute	http://ifs.plants.ox.ac.uk/OFI/OFI.HTM
RAN	Rainforest Action Network	http://www.ran.org/
RIC	Rainforest Information Centre	http://www.ran.org/info_center/index.html
TFP	The Forest Partnership	http://www.together.net/~wow/Index.htm
UN	United Nations	http://www.un.org/
UNEP	United Nations Environment Programme	http://www.unep.org/
USDA	United States Department of Agriculture	http://www.usda.gov/
USFS	United States Forest Service	http://www.fs.fed.us/
WCMC	World Conservation Monitoring Centre	http://www.wcmc.org.uk/
WFI	World Forest Institute	http://www.vpm.com/wfi/
WB	The World Bank	http://www.worldbank.org/
WRI	World Resources Institute	http://www.wri.org/
WTN	World Timber Network, by Lejé International	http://www.transport.com/~leje/wtn.html
WWF	World Wide Fund for Nature	http://www.panda.org/



Internet in the Indian context

Sizigindi Subba Rao

Central Leather Research Institute, India

Abstract: Presents briefly the concept of the Internet and lists the Internet service providers in India viz., Education and Research Network from Department of Electronics, National Informatics Network from National Informatics Centre, Gateway Internet Access Service from Videsh Sanchar Nigam Limited, and SOFTNET from Software Technology Parks India, with their objectives, services and tariff.

Presents the preparedness of India to join the information superhighway by mentioning the activities of various organisations, grouping them broadly under business, directories, investment, newspapers, online information, religious, tourism, exhibitions/conferences, software products and campaign. Points out the limitations to Internet access in India by way of infrastructure with limitations on leased lines, expensive hourly usage costs due to lack of tie-ups with Internet companies and gauging the popularity of Web sites. Concludes that the booming Indian IT market trend encourages the Internet service providers to expand and Internet awareness is on the rise through various seminars/workshops coupled with live demos. In spite of hurdles, India is gearing up to join the Internet bandwagon not only to reap the benefits but also to make available resources and use the Net effectively in sectors like banking, finance, trade, edutainment, healthcare, etc.

Keywords: Internet, Internet service providers, ERNET, NICNET, GIAS, SOFTNET, Indian experiences

1. Introduction

The Internet is the catch-all word used to describe the loose connection of millions of computers at thousands of sites throughout the world, whose users can share information or computer files with no central location or administration of the system. It is a logical, global metanetwork with a virtual information store, arising out of the best instance of a cooperative movement; it transcends political boundaries and is a global arena for hack-ticks. There are companies that help manage different parts of the networks that tie everything together. The networks within different countries are being funded and managed locally according to local policies. Access to the Internet usually means that one has access to a number of services such as electronic mail, listservs, archive servers, netnews, FTP, telnet, gopher, WAIS, WWW and Web authoring. Around the world people make information available and are responsible for the documents they write and make public on the Web. The five areas in which Internet applications would have an impact are sales and marketing, human resources, customer services, engineering/research and development, and financial.

2. Internet service providers in India

The Internet facility, once available only to government, academic and research communities, is now accessible to anyone who has the time and passion for information. The early networks were mainly based on terrestrial lines whereas the recent ones are in satellite mode. The major Internet service providers in India are ERNET (based on terrestrial lines and VSATs), NICNET (mainly satellite based), GIAS (based on terrestrial lines and satellite) and SOFTNET (mainly satellite based).

2.1. Education and Research Network (ERNET)

The ERNET was initiated in November 1986, with the financial assistance of Rs. 10 crore (US \$3 million) from the Government of India and US \$6 million from the United Nations Development Programme. It involved eight premier institutions in the country as participating agencies: the five Indian Institutes of Technologies; the Indian Institute of Science, Bangalore; the National Centre for Software Technology (NCST), Mumbai; and the Department of Electronics, New Delhi.

ERNET was the first network in India that provided access to the global Internet. The user base of ERNET has grown to more than 50,000 users in 600 organisations and provides access to about 120 networks in other countries. ERNET offers the whole range of Internet services through the eight backbone nodes at geographically spread locations. The main sites are being interconnected through terrestrial links and the access to the international gateway is presently through a 64 Kbps link. With the successful commissioning of Sat-WAN Hub, ERNET

sites are being accessed progressively through the VSAT network. The network has two overseas gateways situated at NCST, Mumbai and the Software Technology Park (STP), Bangalore that connect ERNET users to global networks. These gateways link to two local networks in the US and are helping Indian scientists log in to an innumerable number of host databanks that are available on Internet.

ERNET offers three types of access options:

- (1) access through a dial-up telephone line (local or STD);
- (2) access through a leased line;
- (3) access through VSAT (Ref 1).

Subscription to ERNET is on a per annum basis: for dial-up access, terminal type or Serial Line Interface Protocol (SLIP)/Point-to-Point (PPP), up to 200 Kbps of mail traffic per day would be Rs. 1 lakh (US \$2900). With leased line access, up to 64 Kbps and higher volume of traffic up to 1 Mb per day would be Rs. 2 lakh (US \$5700), excluding the leased line charges. The VSAT based link would cost Rs. 2 lakh (excluding the cost of VSAT equipment). E-mail access up to 200 Kbps per day would be Rs. 25000 (US \$700) (Ref 2). Some of the future areas of work include test bed for high speed networking and support of applications like multi-site video conferencing, and other integrated services like multimedia mail and multimedia document retrieval by upgrading its bandwidth to 2 Mbps. The network hopes to serve 10,000 institutions by 1997.

2.2. National Informatics Network (NICNET)

NICNET is one of the extensive nationwide networks that is mainly dependent upon the satellite based data communication of the National Informatics Centre, which is exclusively for government organisations and in operation since 1988. It consists of a master Earth station connected to a host computer at New Delhi. The micro Earth stations — around 700 — are being located at all regional, state, district centres and selected commercial centres, hence providing widest reach in the country (Ref 3).

NICNET maintains its leading edge with the incorporation of a powerful Ku-band based National Information Highway as an overlay network on the existing SSMA/CDMA architecture (comprising VSAT links operating at 1200 bps to 9600 bps). This overlay network is established in 14 cities and is being extended to 70 cities with the central hub at New Delhi. It supports high speed communications up to 2 Mbps. It is being connected to over 200 international networks in 160 countries and has dedicated Internet access through a direct high speed link to Sprintnet, US. The National Information Highway allows its users multimedia Internet access for electronic commerce, distance education, research document support, distributed geographic information system, a legal information system and medical databases among others, worldwide. NICNET established the first WWW server in India: the BASIS Web Server in collaboration with Information Dimensions Inc. added a new dimension to the existing EDI services and bibliographic services on NICNET (Ref 4). NICNET provides Internet access in three ways:

- (1) mail only access;
- (2) shell account — dial-up terminal access;
- (3) SLIP/PPP — on-demand direct access.

The user base of NICNET, apart from the government organisations, is around 2000 and NICNET hopes to raise it to 10,000 by the year end.

2.3. Gateway Internet Access Service (GIAS)

Videsh Sanchar Nigam Limited (VSNL), India's international telecom carrier, has provided full-featured Internet service from its GIAS since 15 August 1995 by connecting its main Internet access node at Mumbai to the Internet node in the US through satellite media and to Europe through submarine cables. It has remote Internet access nodes at Calcutta, Madras, New Delhi, Pune and Bangalore and these are being connected to the Mumbai node through the Department of Telecom's (DOT) inter-city links. The GIAS gives users a full Internet access through a normal telecom services. The user base of VSNL is around 12,000 which includes individual, corporate, and educational institutions.

VSNL provides a platform for the service and offers three types of accounts: (i) Student accounts, (ii) shell accounts and (iii) TCP/IP accounts. Further, VSNL has enriched the services to enable users to create their home pages to make public on the Net, thus enabling Indian companies to be visible worldwide in the cyberspace. VSNL has fixed the tariff on per annum basis and Rs. 5000 (US \$150) for 250 hours of usage for individuals using the shell account with dial-up access. For TCP/IP access the tariff is Rs. 15000 (US \$450). The tariff for commercial organisations is Rs. 25000 (US \$750) for both types of access. For leased line access, the tariff varies from Rs. 1.5 lakh (US \$4300) for commercial organisations using 2.4 Kbps to Rs. 25 lakh (US \$72000) at 128 Kbps. For students the special tariff is Rs. 500 (US \$15) (Ref 5).

All the Indian Internet service providers — ERNET, NICNET and VSNL — are accessing the Internet through the single gateway provided by VSNL. To expand VSNL's Internet operations, it has plans to have a nationwide backbone for smooth flow of denser traffic by enhancing the bandwidth to 2 Mbps and making India a regional hub of connectivity for Asia by linking to Singnet, a Singapore based ISP. It expects a 1 million user base by the turn of the century (Ref 6).

2.4. SOFTNET

Set up by the Software Technology Parks of India (STPI), in association with SatComm Services India, the just-started SOFTNET promises maximum up-time and robust access to software exporters to communicate with their overseas clients without going through VSNL. All connectivity to access points is through radio links that are very reliable. It offers high speed data communication links and direct Internet access. Each of the six STPs at Noida, Bangalore, Bhubaneswar, Hyderabad, Tiruvananthapuram and Gandhinagar has an international gateway.

Softnet offers three types of services:

- (1) SoftPoint, a dedicated leased line service ranging from 9.6 Kbps to 2 Mbps offers interactive applications, video conferencing and large business operations with specific clients or offices abroad;
- (2) SoftLink, a shared Internet service using TCP/IP and designed to provide both the IP routing service as well as server-based access with facility for virtual hosting;
- (3) SoftPack, an X.25 based packet switching network service that allows users to access other international public data networks.

Services like video-on-demand, bandwidth on demand and other value added services viz., authoring of Web pages and setting up Web servers for providing single point service are on the way (Ref 7).

SOFTNET tariff is on a per annum basis. For dial-up mode, SLIP/PPP (packet billing), V.34 modems, 16-32 lines would be Rs. 26000 (US \$750) plus Rs. 150 (US \$5) per Kilo segment (64 K Bytes). For dial-up access, flat fee billing, V.34 modems, 16-32 lines would be Rs. 4 lakh (US \$11500). The radio link 64 Kbps access would be Rs. 5 lakh (US \$14300) per 64 Kbps plus radio link charges Rs. 1 lakh (US \$3000). The LAN access within STP complexes would be Rs. 4 lakh with 64 Kbps (Ref 2). With the basic network in place, next on the agenda are plans to upgrade the level of bandwidth radiated from the Earth station for SoftPoint and SoftLink services to 2048 Kbps for the US, 1408 Kbps for European countries and 256 Kbps for Asia Pacific.

2.5. Corporate agencies

Apart from the four government funded ISPs, two more are on the way from corporate agencies, viz.:

- (1) *InterNet India*: a dial-up Internet service launched by J.H. Communication Pvt Ltd., Hyderabad in May 1996, targeted at corporate clients for their high speed and reliable Internet access, with working cost of less than one rupee (US \$0.02) per kilobyte of data transfer;
- (2) *Computer Maintenance Corporation Limited*: the CMC Limited has tied up with VSNL and Pritish Nandy Communications to provide Internet access, both in India and abroad, to establish and run Internet sites and 'Cybercity' at Mumbai. These sites would provide news and an interactive databank of corporate information, besides offering information on India's culture, heritage, education and entertainment.

3. Indian experiences

The Web sites that came up have originated mainly from corporate agencies, with servers located in US. Several others are in the pipe line. The ones that are from government agencies are mainly focusing on the investment, tourism and edutainment. The Web sites are broadly grouped under Business, Directories, Investment, Newspapers, Online Information, Religious, Tourism, Exhibitions/Conferences, Software Products, and Campaign.

3.1. Business

Following are some of the Web sites that cover business information on their servers.

- (1) *IndiaWorld* (<http://www.indiaworld.com/rdc>): the first Web site launched in March 1995, from Ravi Database Consultants Pvt Ltd., Mumbai with Indian interests along with ASAP Solutions Inc., US. It offers a range of news and information, business, finance and technology from magazines such as *India Today*, *Business India* and *Express Investment Week*. One year's subscription to IndiaWorld costs Rs. 700 (US \$20);
- (2) *Rediff on the Net*: the first interactive online service on the Internet from India launched by Rediff Communications during February 1996. The server was housed at Mumbai and has a mirror site at California. It covers a wide range of domestic and international topics such as current events and news, business information, personal investment guidance, films and entertainment information, travel and leisure tips, sports coverage, specialised sections for students and children, and a comprehensive archival database. The subscription charge for the free to view area is Rs. 1600 (US \$45) per annum and for the complete service is Rs. 4800 (US \$140) per annum;
- (3) *G On Line* (<http://www.chitralekha.com>): the first online film magazine from Chitralekha Group, Mumbai launched in January 1996, targeting the NRIs. The site offers films, finance, politics, some free services, additional information with subscription; pay per use service and advertising;
- (4) *Gateway of India*: launched by Sound Electronics, Mumbai to install the pages on the Net and gained

popularity with major companies opting for advertising on it;

- (5) *Made in India*: a Web site from the Mumbai based ad agency Folklore Communication Management Services Ltd, started Internet advertising and dedicated to exporters of products and services from India, providing them with opportunities for online advertisement and promotions;
- (6) *Connect! India*: launched by Elnet Technologies, Madras with the Web site at Virginia. It is a virtual business mall on the WWW, dedicated exclusively to Indian business. Every client on this mall is provided with a host of services ranging from design, creation and hosting of home pages, taking care of publicity, feedback, registration of address and search tools;
- (7) *Sabre Opportune*: launched during February 1996 by Strategic Advantage for Brand Equity, a Bangalore based advertising agency advertising its services on the Internet, targeted for the organisations interested in India as a market. It provides an overview of India as a marketplace, the category of product or service profiles of the intended customers and recommend possible niches;
- (8) *Info-Shop*: floated by The Jobnet Ltd, Bangalore during April 1996 to shorten the distance between searching and finding jobs, limiting the need to go through the classified columns. Jobnet would cover all kinds of jobs on offer and intends going to the US too, apart from South East Asia;
- (9) *India Trade* (<http://www.indiatrade.com/cerc>): the Consumer Education and Research Centre, in collaboration with its Atlanta associate, publishes a business magazine *India Trade*.

3.2. Directories

The directories that are available on the Internet include:

- (1) *SICOD'96*: the first Indian commercial directory, *Southern India Commercial Directory '96* from Metro Information Services Pvt Ltd., Madras lists more than 12,000 companies in South India with essential information on the company and its products (Ref 8);
- (2) *Exporter's Directory* (<http://www.leatherindia.com>): from the Council for Leather Exports, Madras to host electronic exporters of Indian leather and leather products directory on the Web. One listing would contain the company profile, its products and other basic details and the other would have home pages for whole product catalogues.

3.3. Investment

The following Web sites cover investment related information.

- (1) *Reserve Bank of India* (<http://www.indiaworld.com/home/rbi>): RBI has reserved three home pages on the Internet for disseminating banking and foreign exchange information;
- (2) *Business Information Services Network (BISNET)*: the Federation of Indian Chambers of Commerce and Industry is going on the Internet with a huge database on the Indian economy and business, with daily updating and allowing the browser to go through the 150,000 pages;
- (3) *IPAnet*: launched on 15 March by Multilateral Investment Guarantee Agency (MIGA), the World Bank's insurance arm. This electronic investment mall contains global databases on foreign investment conditions, from the mining laws prevalent in each country to the investment codes to the business operating conditions. MIGA plans to open IPAnet free of cost for six months and later on will be pricing at US \$15 per month;
- (4) *National Stock Exchange (NSE)*, Mumbai is being prepared to go live on the Net to aid investors, especially NRIs for placing orders while keeping a tab on real time stock prices;
- (5) *Securities and Exchange Board of India (SEBI)*, Mumbai will be going online in three months with a home page of its own, aiming to bundle all its regulations backed up with regular updates, data on changes and events in the capital markets and annual reports (Ref 9);
- (6) *EEPC site* (<http://www.eepcnet.com>): the Engineering Export Promotion Council has plans to provide information about the Indian engineering export community on trade statistics, market intelligence, joint ventures and international regulations.

3.4. Newspapers

The three Indian news papers that are available on Internet are:

- (1) *The Hindu: The Hindu*, one of the national newspapers went online on the Internet as the *Hindu Online (HOL)* to cater for the needs of overseas readers during April 1995 and became a weekly edition from the first week of June 1995. About 350 items of news/articles/features along with images, classified under 30 categories posted on a server in the US with the cooperation of Paralogic Corporation. The site received its first recognition as one of the Top Five Percent sites by Point Communications Inc. A daily service with the cooperation of the India Information Inc. was launched on 1 October 1995 on a different server, Indiaserver in New York. This site provides news briefs and business and investment related information, and bagged the Gold Tiger Award given by Worldclass.

- *Hindu Online*: <http://www.webpage.com/hindu/index.html>;
 - *The Hindu*: <http://www.indiaserver.com/news/thehindu/thehindu.html>;
 - *Business Line*: <http://www.indiaserver.com/news/bline/bline.html>.
- (2) *Indian Express* (<http://express.indiaworld.com>): *Indian Express* was one of the nine sites on the Internet that offered news about Parliamentary elections in India during April 1996. Over one million hits made and about 300,000 documents downloaded by NRIs from 75 countries at the Express India World Election site, which launched when the polls began. On the Net, the election site divided into three categories: national results with analysis and graphics; a state-wise result chart with background and graphs; and a constituency-wise chart. The group has plans to launch its publications *Indian Express* and *Financial Express* on the Net;
- (3) *The Economic Times* (<http://www.economictimes.com>): one of the sites that offers detailed financial and corporate information from Bennett, Coleman & Co., Ltd., Mumbai.

3.5. Online information

The sites that offer information through online include:

- (1) *Medical*: the medical library at the Malpani Nursing Home, Mumbai extensively searches the Paper Chase Library (<http://www.pch.bih.harvard.edu>) on the Internet. To make it more economical the library has tied up with Beth Israel Hospital, Harvard to provide access to the Cyber library for Indian doctors by doling out passwords for a non-profit fee Rs. 2000 (US \$60) (Ref 10);
- (2) *ERL Technology*: Global Informations Systems Technology Ltd., (GIST), New Delhi has unleashed Electronic Reference Library (ERL) technology in India, from SilverPlatter Inc.;
- (3) *Indian Union Budget* (<http://www.nic.in/indiabudget>): for the first time the union budget for 1996-97 was made available covering economic survey, budget at a glance, highlights, speech, customs and central excise notifications and annual financial statement.

3.6. Religious

The sites that cover Indian religion are:

- (1) *Bhagavad Gita*: The Centre for Development of Advanced Computing, Pune has undertaken a project 'Dnyaneshwari', the famous treatise on the Bhagavad Gita. The volume of work would be around two million characters and translated into English for making public over the WWW (Ref 11);
- (2) *Ramakrishna Mission* (<http://www.daistech.com/rkmission>): DAIS Infotech hosted a Web page for Ramakrishna Mission, the socio-religious organisation, covering its origin, various centres, activities and services, with 'Belur Math' as its focal point.

3.7. Tourism

The Web sites that cover tourism information include:

- (1) *Mumbai on the Net*: provides up-to-date and authentic information about Mumbai city. It was launched on 22 May 1996, sponsored by the Maharashtra Tourism Development Corporation;
- (2) *Gujarat*: the Tourism Corporation of Gujarat Ltd. is to woo domestic and international tourists by providing general information, pictures, and details of places of historical and cultural importance of Gujarat by setting up a home page on the 'Royal Oriental Express' (Ref 12).

3.8. Exhibitions/conferences

The Web sites that held exhibitions/conferences include:

- (1) *NASSCOM'96* (<http://www.nasscom.net.in>): the National Association of Software and Service Companies launched the first live server on the Internet from India to cyberspace the Nasscom'96 conference and exhibition organised in Mumbai during 7-10 February 1996, and provided updated information on the conference and exhibition;
- (2) *PCL Mindware* is gearing up to launch a 'virtual art gallery' to surf across the Internet with more than 1000 paintings and the entire range of Indian music, and functions as an Internet designer to assist Indian corporates in hosting their home pages on the Web. It also caters in the areas of corporate presentations, edutainment, kiosks and computer based trainers on the multimedia front.

3.9. Software products

The Web sites hosting the software products include:

- (1) *CAD Products*: the Madras based California Software Ltd launched its CAD suite of products on the Internet for distribution. These include a set of application programming interfaces for developers of CAD applications, a custom control to view drawings (created in AutoCAD) that could be embedded in office documents such as Word and Excel and presentation tools like Power Point, and an enterprise

management system for fuel oil bunkering;

- (2) *Fact*: The Internet version of an accounting applications software from Vedika Software, Calcutta. Its Internet server version would sit on an Internet server anywhere in the world and the client front end interface would be able to access the package on the server, and yet register transactions and pull out end use data as if from a local site. The company accesses the WWW through leased lines of the Internet Research and Development Unit of the National University of Singapore and has an interactive home page (Ref 13);
- (3) *Millennium Masala*: an electronic magazine on the Internet focused on young professional NRIs in North America and UK by Millennium Cyberworks. Masala offers Indian themes and addresses the Indian market with Internet solutions while offering Internet development skills to the Indian workforce in creating and marketing their Web sites. To assist Indian businesses it has created SAFFORN (South Asian Firms, Finance and Realty Online), a virtual trade show and directory on the Web for Indian export houses, financial institutions, real-estate developers and individuals to reach western buyers, partners, investors and employers.

3.10. Campaign

The following groups are devoted to campaigning for the Internet:

- (1) *India Internet Java Users Group (IJUG)*: formed after the release of Java during December 1995 to campaign intensively for the cause of Java by way of seminars, training programmes and interactive demos; to demonstrate the power of Java programming tools and techniques; to serve as the world's first Java Resource Centre; to interact with Java users the world over; and to keep Indian IT professionals up-to-date with application development techniques (Ref 14);
- (2) *Silicon Graphics India*: hosting a series of seminars in India to highlight its strengths in the Internet and with Netscape Communications to educate users about the Internet, and the latest Web authoring and serving techniques;
- (3) *Cyberskool*: a high-tech multimedia computer lab 'Cyberskool' with Internet access was supported by Intel's Corporation at the National Science Centre, New Delhi;
- (4) *Cyberclub*: India's first Cyber Cafe that started during 1995 at Maurya Sheraton, New Delhi evolved into a serious Cyberclub. Individual membership costs Rs. 30000 (US \$900) and a corporate membership Rs. 50000 (US \$1500) annually, with browsing hours limited to seven hours a week.

4. Limitations

The Internet, though a very common thing in the West, has to clear quite a few hurdles before the average Indian can don the interning cyber suit. Existing limitations include:

- (1) *the infrastructure*: India has one of the lowest telephone penetrations in the world with 1.3 per 100 as compared to the world average of 11 telephones per 100. Currently Internet access is only through VSNL. Though the private companies get the nod to provide access, they have to rely on VSNL for the backend. As access is through telephone lines, there is need for a higher telecom efficiency than India currently has;
- (2) *costs*: VSNL currently offers shell and TCP/IP accounts. The costs are high even for corporates and it works out to be around Rs. 100 (US \$3) an hour plus the STD call charges where the subscriber does not have direct access to a local node.

These limitations have certain implications on advertisers in India:

- (1) limitations on leased lines in India: downloading of graphics takes a long time and one has to pay for the phone calls due to lack of tie-ups with Internet companies;
- (2) Internet marketing and advertising generally converts into a purchase by just punching in the credit card number or by informing the bank. This seems impossible in India as the tendency in the country is still to verify the signature of the credit card holder on the pay slip;
- (3) for the advertiser, gauging the Web site's popularity based on the number of hits could be misleading. However, this problem is not just limited to India but cuts across the Internet spectrum across the world.

5. Conclusion

The IT industry in India during 1995 showed a compounded growth rate of 50% (over Rs. 6000 crore, equal to US \$1.7 billions) with 300,000 PCs installed (60% growth rate), applications software segment 100% and peripherals with 400%. Reduction of import duty on software to 10% and on computers and peripherals to 40% encourages the SOHO and home markets to pick up soon. The financial institutions and the banking sector are the current areas of adoption of IT in India. The launch of an Internet service by VSNL during August 1995

attracted much enthusiasm and reminded of the need to gear the telecom infrastructure to address the needs of the volume of traffic and transactions that are needed for commerce over the Net. The DOT claims that the backbone is in place: 80% of the main trunk lines are already fibre that link the country's major cities with 140 Mb, to expand to 500 Mb. The privatisation of basic telecom policy by the government also allows many corporate organisations to jump into the Internet arena for healthy competition. Awareness about the Internet is on the rise is the past year through various workshops, coupled with live demos spanning from one day to three day programmes by academic and commercial organisations, held on average once a month throughout the country. In spite of hurdles, India is gearing up to join the Internet bandwagon, not only to reap the benefits but also to make available resources on the Net. India could also effectively use this information highway in sectors like banking, finance, trade, edutainment, healthcare and so on.

S. Subba Rao
Scientist, Information Science
Central Leather Research Institute, Adyar, Madras - 600 020, India
E-mail: ssrao@niclai.ernet.in

References

- [1] Department of Electronics (1995) Education and Research Network brochure.
- [2] Ghosh, A. (1966) ERNET and SOFTNET: Internet connectivity for target groups, *PC Magazine India*, **28**(2), 33-38.
- [3] Singh, S. (1995) Data communication network: remove the speed barriers, *The Economic Times*, Madras, 3 May, 7.
- [4] National Informatics Centre (1995) World Wide Web for Internet Services through NICNET Info Highway brochure
- [5] Videsh Sanchar Nigam Ltd. (1996) Gateway Internet Access Services brochure
- [6] Kumar, B. (1996) Many happy returns of the day, *The Economic Times*, 19 August, 6.
- [7] Verma, P. (1996) SOFTNET: one more gateway, *The Economic Times*, 6 May, 6.
- [8] Metro Information Services (1996) SICOD'96 brochure.
- [9] Ethiraj, G. (1996) Sebi to join Internet bandwagon, *The Economic Times*, 12 June, 8.
- [10] Kumar, B. (1996) Optimum information intake for medicos and public, *The Economic Times*, 2 April, 6.
- [11] Joshi, H. (1995) Bhagavad Gita treatise ..., *The Economic Times*, 19 December, 7.
- [12] Krishnadas, K.C. (1996) See Gujarat, on the Internet, *The Economic Times*, 7 July, 3.
- [13] Bakshi, J.V. (1996) Vedica announces software for Internet, *The Economic Times*, 4 January, 15.
- [14] Kumar, B. (1995) Java users group ..., *The Economic Times*, 19 December, 17.



Looking for the information needle in the Internet haystack

Helge Clausen

State and University Library, Denmark

Abstract: In the present paper the most important findings from a comprehensive empirical survey among 413 Danish Internet users are presented. The survey was carried out in the winter of 1995-96. Special attention is paid to user patterns and information quality. Differences in expectations among various groups of users (business, higher education/research and students) are analysed and discussed. Comparisons with findings from other relevant surveys are made. On the basis of some of the empirical findings, two models concerning the placement of Internet information resources are discussed.

Keywords: Internet, information quality, Denmark, research libraries, librarians, information professionals, controlled information network environment, virtuality, World Wide Web

1. Introduction

The information resources accessible through the Internet continue to grow in extent. An increasing number of users are trying to satisfy their information needs by using such resources. Are they getting stones for bread?

According to a recent survey from the Gartner Group (Ref 1) about half of the world's 280 million PCs will be connected to the Internet by the beginning of 1997. That is the optimistic forecast, while a more conservative one — also from the Gartner Group — says 10%. That means that between 28 million and 140 million PCs in a few month's time will become interconnected, and be potential sites for the input and retrieval of all kinds of information. It is thought-provoking, however, that for example 100 million people form less than 2% of the world's population. Even today, a century after the invention of the telephone, about half of mankind still does not have access to it.

The Matrix Information and Directory Services (MIDS) distinguishes between three concentric entities in calculating the number of Internet users (Ref 2): the *Matrix* includes the *Consumer Internet* which includes the *Core Internet*:

- *Core Internet*: 16.9 million users of 7.7 million computers that can distribute information by interactive TCP/IP services such as the World Wide Web (WWW) or FTP;
- *Consumer Internet*: 26.4 million users of 10.1 million computers that can access information by interactive TCP/IP services such as WWW or FTP;
- *The Matrix*: 39 million users of electronic mail.

The figures are taken from the latest MIDS Internet Demographic survey from October 1995 (issued in February 1996). This way of making up the number of Internet users gives a rather conservative estimate that lies between the optimistic and conservative forecasts of the Gartner Group mentioned above. Further, MIDS has calculated that the Internet has been growing at about 100% per year since 1988.

How can the Internet-based resources be seen? A Danish futurologist has pointed out that the Internet is not being steered because that is impossible. One can't steer or manage a jungle. The Internet is pure Darwinism, which means that everybody does as one likes and in the long run it is about 'survival of the fittest'. Now that the Internet has been commercialised, a few years ago a more exact phrase would be 'survival of the profitfittest'!

A balanced meta-point of view has been formulated by Floridi (Ref 3, p. 45):

'The global network is only a stage in the endless self-regulating process through which the human encyclopedia constantly strives to respond to its own growth.'

Let us try to be modern encyclopedists, who take nothing for granted but who are willing to see visions and dream dreams.

2. Empirical findings

2.1. The surveys

During the last two years the present author has carried out studies of the Internet as a tool for research libraries. The studies have been supported by the Danish Ministry of Culture from which the author received a £12,000 grant. Besides following the developments of the Internet as a user, the author has planned and carried out several studies of its use by Danish information professionals and common users. In 1995 a pilot study and a forecast (Delphi study) were completed. The most interesting of the findings have been published (Ref 4, pp. 65-67; Ref 5).

The general interest in figures regarding the Internet decreases in an age of an increasing number of users. Who is interested in the number of radio listeners or telephone users nowadays? In the Western World almost everyone has a radio and a telephone, among many other things. In five years or so from now a majority of the same people will also have access to the Internet as a result of computerisation and cheaper telecom networks.

Below, findings from two different Danish surveys will be presented.

2.1.1. Methodological problems — for safety's sake.

Many Internet surveys are being carried out by asking users to complete an electronic questionnaire and return it by e-mail. In the nature of the case such surveys cannot be claimed to be representative of the population that is being investigated because respondents only make up a random sample. A strictly representative survey would be almost impossible to carry out because the whole group of Internet users in a country or of a certain category is not known, or is described by means of quantitative terms. These circumstances may call the reliability of the survey into question. As matters stand, one can only lean against the findings from other surveys which are carried out in different ways, e.g. of the Gallup poll type. It should however be kept in mind that many Internet surveys are being carried out by asking random users to complete an electronic questionnaire, e.g. GVU's WWW User Surveys (see below). Finally, according to the Law of Large Numbers, when the number of respondents increases, so too does the probability that they make a representative sample.

2.1.2. The author's survey.

In the winter of 1995-96 the present author performed a comprehensive empirical survey in order to throw light on Internet user behaviour. Three special groups of users (business people, higher education/research and students) were selected for further analysis. Focus has been on the use of the Internet as a communication tool and as a means of utilising information sources. A questionnaire was sent to all customers of four of the largest Internet distributors in Denmark (UNI-C, IBM Denmark, Cybernet, and Jyllands-Posten). The distributors agreed to support the survey by distributing the questionnaire via e-mail or traditional mail to their Internet subscribers, the number of which was confidential. A total of 413 filled-in questionnaires were returned, 97% of them via e-mail. The respondents had been Internet users for 18.4 months on the average. Two thirds of the respondents came under the three special groups selected for further analysis: business people 35.4%; higher education/research 19.2%; students 11.6%. Forty-seven percent of the respondents had complied with the author's request to add comments.

2.1.3. The Mediacenter Danmark Internet survey.

Data from a recent Danish Internet survey are also included. The survey was performed in March 1996 for Mediacenter Danmark by Gallup Denmark, and builds on telephone interviews with a sample of 2981 persons (Ref 6). The sample was made by a random selection of telephone numbers. Only persons over 13 years of age were interviewed.

2.2. The users — who are they?

According to the Mediacenter survey (Ref 6, pp. 4, 17), 11% of the adult Danish population has access to the Internet but only 7% are actually using it (10% of all men and 4% of all women).

In the author's survey the following distribution of age for respondents was found (Figure 1).

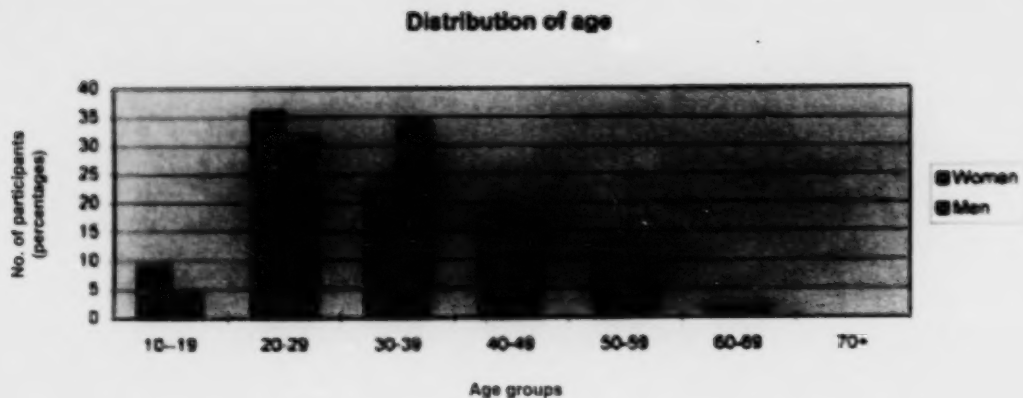


Figure 1: The author's survey: Internet user survey. Distribution of age for women and men (percentages).

There is no doubt that people between 20 and 39 years of age dominate the group of respondents. The average age is 35 (ranging from 15 to 71). Female respondents are in the minority but from the relative figures they seem to be in the majority in all age groups, except for 30-39. In the Mediacycenter survey the age group for 25-39 makes up 37% of the respondents and the group for 40-59 another 32% (Ref 6, p. 5).

From a comparison with six similar findings from 1995 it must be concluded that the percentage of female Internet users is 30-40% and is increasing (Ref 7):

'The Internet has traditionally been dominated by males, but this is changing. As more studies of the Net are conducted and research methodologies improve, female users continue to account for greater percentages of the online community. These results should provide valuable fodder for online commerce.'

At the Graphic, Visualization & Usability Center (GVU) at Georgia Technical Research Corporation (Atlanta, Georgia) WWW user surveys are being carried out regularly (Ref 8). The fifth and latest survey (April-May 1996) includes 11,700 respondents, of whom 31.5% are reported to be female, compared with 29.3% for the fourth survey half a year earlier and with 15.5% for the third survey (April 1995). A comprehensive survey of Internet users by A.C. Nielsen (referred to in Ref 9) includes 34% women.

These indications of an increasing female use of the Internet should be compared with the way women are using it (see Section 2.3.).

Further, a recent American survey made by the Online Research Group has shown that Internet users form a younger group than those of other online services (Ref 10). It is understandable that the young users are attracted to the more exciting colours, graphics and variety of services of the World Wide Web. The so-called 'digital kids' will become the future Internet users, and marketers have already begun to focus on the millions of children who are Internet users. In 1995 an estimated three million children in the US had access and it is estimated that 12 million children will have access by the year 2000 (Ref 11, p. 45).

2.3. Internet use — how much and by whom?

An important factor is the period of time the user has been occupied with the Internet. The respondents in the author's survey had been users for 18.4 months on the average. Sixty-eight percent of the respondents had 12 months' experience or less. The average for women was 15.3 months and for men 19.1 months. Concerning the size of current use, the respondents are connected to the Internet 6.14 hours/week on the average. Women and men, however, seem to have different patterns of Internet use. In a comprehensive survey carried out by A.C. Nielsen (referred to in Ref 9) the average Internet user is connected 5.47 hours/week.

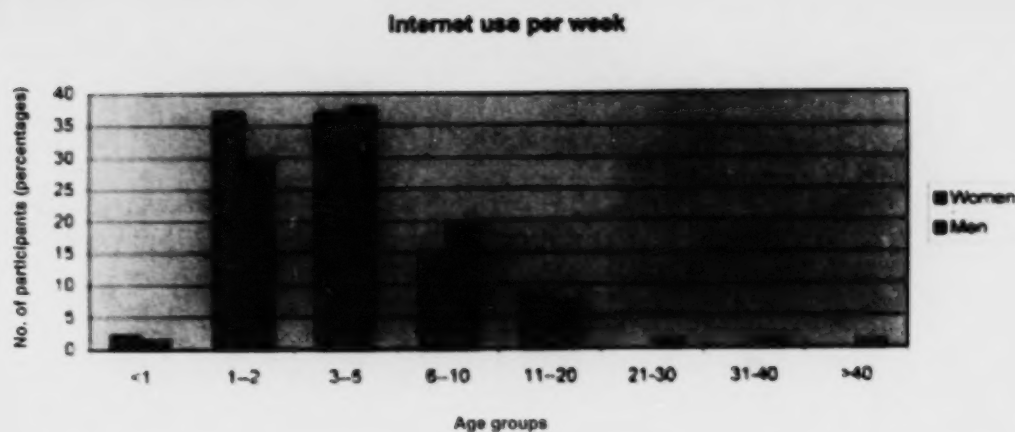


Figure 2: The author's survey: Internet use in hours per week for women and men (percentages).

Figure 2 shows that women are in the majority in the group as people who use it less than three hours per week, while men in general are connected for longer periods of time. On the average women use the Internet 4.79 hours/week and men 6.35 hours/week. This is comparable with a result from an earlier pilot study where it was found that Danish information professionals used the Internet for 4.1 hours/week on average in May 1995 (Ref 4, p. 65).

The Mediacenter survey has some findings that support the hypothesis of women using the Internet for shorter periods of time than DO men (Ref 6, p. 25):

Table 1: Duration of Internet sessions in minutes: Distribution of women and men (Mediacenter survey).

	1-5	6-15	16-30	31-60	60+
Women	23%	21%	28%	16%	11%
Men	16%	22%	27%	22%	13%

Concerning sessions lasting 1-5 minutes, women are in the majority while men clearly prefer longer sessions (more than 30 minutes).

The above findings concerning female and male Internet use are in harmony with several comments from the respondents of the author's survey stating that women tend to use the Internet to solve well-defined problems, while men are inclined to do some net surfing (called 'net watching' by the serious users!) as well. When it comes to information retrieval, female Internet users seems to be more goal-directed and convergent. In most cases they stop when the information searched for has been found. Male Internet users, on the other hand, have adopted a more divergent/explorative attitude which may lead to experiences of the serendipity¹ type once the information searched for has been found.

In the Mediacenter survey, net surfing is the most popular Internet activity followed by database searching and e-mail in the second and third places (Ref 6, p. 18).

2.4. Electronic mail — If I send a letter to you what will happen?

The classic use of the Internet is communicating by sending and receiving electronic mail, including non-textual files. In the author's survey the average user sends 13.3 messages/week (ranging from 0-170) and receives 21.3 messages/week (ranging from 0-1000). The trend is that users who send up to 20 messages/week receive fewer than they send. In the case of over 20 messages/week sent, more messages are received than sent. Thus, there seems to be a 'critical mass' of electronic mailing. Over 20 messages/week could bring about snowballing! And that might happen in step with an increasing use of, for example, distance work/telecommuting.

According to a survey from October 1995 (Ref 12), 20% of US workers will be telecommuters in 2003. Such findings should always be taken with a grain of salt. In the author's survey respondents reported a total of 13.2% of their working hours as distance work. In five years the figure was estimated to a total of 24.4%, which means an increase of 85%, or 17% annually on the average. So far, things have not developed that fast concerning distance work.

2.5. Electronic conferences

Another popular Internet facility is participating in electronic conferences/discussion groups. The respondents in the author's survey subscribed to 5.2 electronic conferences on the average. Fifty-nine per cent, however, did not subscribe at all and 24.3% had limited themselves to fewer than six conferences, which seems to be the maximum for most users provided that regular activities are going on. The information professionals in the above pilot study participated in four conferences on average.

Concerning the number of messages, the respondents only sent 2.7 messages/week (ranging from 0-40) and received 181 messages/week (ranging from 0-3500) on the average. This high number of received messages was mainly due to a few users who gave very high numbers. In the pilot study the information professionals sent 0.8 and received 107 messages/week on average. As was the case in electronic mail, a similar 'critical mass' could be identified in electronic conferencing: Users sending fewer than ten messages/week receive fewer than they send, and users sending more than ten messages/week receive more than they send.

2.6. Quality in information retrieval

A general outline of quality issues in connection with information retrieval can be found in Ref 13, pp. 46-54. The question is further developed in Ref 14.

Most of the respondents in the author's survey were using Internet resources for information retrieval. In their opinion this had improved the quality of their reference work to some extent. On a rating scale from 0 (not at all) to 5 (very), the three special groups had the average scores shown in Table 2.

Table 2: The author's survey: Average ratings of improvement of information retrieval quality for three special groups.

Groups of users	Average rating
Business people	3.65
Researchers	3.46
Students	3.41
All respondents	3.52

Figure 3 shows the different experiences of quality change more clearly. In the business peoples' opinion the improvements have been remarkable, while the researchers and the students are more reserved.



Figure 3: The author's survey: experienced improvement of information retrieval quality (in relation to the situation before use of Internet). Numbers as percentages of respondents.

In the pilot study the information professionals only scored 2.1 on the average. This is quite remarkable compared with the above ratings. A hypothesis could be that increasing information professionalism results in a more sceptical attitude towards the Internet-based resources, their lack of organisation and the primitive search tools available (compared with traditional online and CD-ROM searching). The above findings are not very suitable for a proof or disproof of the hypothesis, but the information professionals at least know where the shoe pinches.

Further, the Internet users were asked to indicate their expectations of the quality of Internet information sources in three years' time. The rating scale used ranged from -5 (considerably worse than today) over 0 (same as today) to +5 (considerably better than today). The three special groups of users had the average ratings shown in Table 3.

Table 3: The author's survey: average ratings of expectations of change of quality of Internet information resources for three special groups.

Groups of users	Average rating
Business people	3.07
Researchers	2.52
Students	2.58

Figure 4 gives the detailed findings of this part of the survey. It is notable that only very few respondents express negative expectations. Further, business people are more optimistic than academics and students.



Figure 4: The author's survey: Expectations of change in quality of Internet information resources for three groups of respondents. Numbers as percentages of respondents.

Students, on the other hand, seem to split into two main groups: the *moderate pessimists* (rating 0) and the *optimists* (rating 5). In this way the students have adopted an attitude similar to the information professionals in the pilot study. At this rating, where the average was 3.0, the same two separate groups could be identified. Elements of an explanation may be found in the fact that students have more contact with information professionals than do the other two groups.

There is no denying that there is a widespread need for quality control of Internet information resources. This need has been formulated in a clear-cut way by Floridi (Ref 15, p. 51) who maintains that:

'Unless academic and cultural institutions provide some form of quality control, we may no longer be able to distinguish between the intellectual space of knowledge and a very polluted environment of junk mail and data.'

This important concern about the very content of the information field leads to a closer analysis of the situation, and an attempt at pointing out some possible themes that should be brought into focus of the information professionals' endeavour to overcome the above problems.

3. Two models for the library information network environment

Why are many people displeased with the Internet as an information source? A recent survey (Ref 16, pp. 136-139) among British business librarians points out some of the most serious disadvantages (Table 4).

Table 4: Disadvantages of the Internet as an information source.

Disadvantage	Importance weighting
Lack of organised information	28%
Lack of quality of information	21%
Lack of relevant information	19%
Lack of reliable information	18%
Too much information	14%
Total	100%

According to the survey business librarians consider Internet information sources *over-hyped*. The biggest concerns are retrieval problems, which are seen as rooted in the legendary and spreading lack of organisation in the Internet. Among further complaints are the lack of authoritative data, disappearing or moving information, and the fluctuating quality of information from site to site. All things considered, 'the Internet is for people who have enough time to browse through irrelevant information in search of pearls. But for business librarians this browsing becomes more like ploughing' (Ref 16, p. 138). The respondents admit (or rather, concede), however, that there are some valuable information sources on the Internet, and many of them have built up their own portfolio of useful sites.

A new phenomenon has come up in reference work in general: patrons are increasingly asking for 'Internet information'. They assume that the information searched for is (only) available on the Internet. Moreover, if the information is actually found on the Internet, they assume that it must possess a number of qualities for that reason alone. The reverse argument has also been stated: if something is not on the Internet, the information simply does not exist (Ref 16, p. 139).

In the author's opinion the above attitudes and trends concerning Internet information sources have the common denominators of:

- *ignorance* of what information literally is (at last, information is not dependent on the media);
- *confusion* of different things (e.g. real time and retrospective databases);
- *Internet superstition* (the glories of the Internet transferred to a single information source).

Many of these difficulties might be prevented if all persons involved would realise that today's information jungle (including the Internet sources) comprises several categories that differ essentially concerning information quality in the widest sense of the word.

A suitable model for putting the record straight is one inspired by L. Costers (Ref 17, pp. 319-320). According to Costers research libraries, in offering electronic services, should guarantee the quality of such services and of the documents provided. The libraries should cooperate in order to create a *Controlled Information Network Environment*. This will give the information professionals an opportunity of selecting Internet-based information resources according to some quality criteria (creating 'virtual reference desks'). Costers allows for three different service layers to develop (Table 5).

Table 5: Costers' model of service layers in libraries.

Service layer	Library services
1	Own electronic services (e.g. OPAC). Services bought to the library (e.g. CD-ROMs, full-text databases).
2	Electronic services of a regional/national library cooperative to which the library contributes.
3	The Internet and its information resources

It is easy to see that in these three different service layers or information network environments, the possibilities of exercising quality control are decreasing in step with development. A library has 100% control over its own OPAC, a little less for, say, a joint national OPAC network, and very little concerning the Internet. Consequently, the library cannot answer for the information quality in the same way in all three layers. Information professionals ought to have a good grasp of these matters but, unfortunately, this is not always the case. Library users may be in an even more unfortunate position by virtue of lacking a general view of the situation. Library users will only be able to make the best possible use of the different information resources if they fully realise the importance and consequences of the different quality of service offered by a particular library.

The concept of 'information' is not an easy one to think of. In order to describe and perhaps even understand

the trends in the development of contemporary and future information systems one also has to use the concept of 'virtuality'. Originating in the subject of computer simulation, this concept has been used increasingly in the information field during the last 6-8 years. The very subtle concept of 'virtuality' (not to mention 'cyberspace') may be even more difficult to understand than 'information', and thus a number of information professional/user misunderstandings are imminent. A usable and operational definition of virtuality is offered by Michel Bauwens, who has contributed extensively to the development of these matters. According to Bauwens (Ref 18, p. 2):

'... virtualisation means that every piece of information produced everywhere in [an] organisation, is accessible from anywhere, anytime.'

When is all this going to happen? It has been on its way since the very beginning of the online age. According to a recent study of the future of libraries, free Net access in American public libraries will be available in 2003; the first virtual large library will open up in 2016; and half of the Library of Congress will be digitised in 2043. The figures are average estimates made by five experts (Ref 19).

Michel Bauwens maintains that virtualisation is going to take place in the libraries during a series of steps which will overlap (see Table 6).

Table 6: Degrees of virtuality in the development of libraries (Bauwens).

Degree of virtuality	Access	Documents
1	Electronic, to OPAC	Paper
2	Electronic, to virtual collections	Paper
3	Electronic, to virtual collections	Electronic

This development will necessarily include:

- the replacement of printed matter by electronic information;
- increasing use of computer-mediated communication;
- universal access to information.

In the author's opinion many pitfalls and talking at cross purposes could be avoided if information professionals and users shared a certain minimum of understanding concerning these different levels of service, virtuality and information quality. After all, this is only a bringing up to date and an extension of some traditional Library and Information Science core concepts. Until the online age librarians often experienced similar problems when patrons confused, say, books with bibliographies.

4. Conclusion

It is an old sociological rule of thumb that technology is moving faster than the institutions, which in turn are changing faster than attitudes.

Anybody knows that information technology is developing very fast at present. In the above case of Internet users and use it was found that some groups of users (women, young users) are running fast in order to keep up with the times and, preferably, to gain on technology. That leaves behind the institutions some of which in the information field might need to rearm in order to survive; especially concerning distance work, which it is believed will develop rapidly in the near future and require new thinking on management, responsibility, and social issues.

A more profound understanding of information quality in the light of virtualisation could lead to less disappointment over the whole lot of the Internet.

Acknowledgements

The author would like to thank the Danish Ministry of Cultural Affairs for the £12,000 grant; and UNI-C, IBM Denmark, Cybernet and Jyllands-Posten who kindly supported his Internet survey.

Dr Helge Clausen
State and University Library
Universitetsparken
DK-8000 Aarhus C
Denmark
Tel: +45 89 46 22 64
Fax: +45 89 46 22 20
Private tel/fax: +45 86 99 63 54
E-mail: hcl@kumsb.dk

References

- [1] Nu skal Internettet tages alvorligt (1996) (Now the Internet should be taken seriously), *Datatid*, 5, 22.
- [2] URL: <http://www2.mids.org/ids3/pr9510.html>
- [3] Floridi, L. (1996) The Internet: which future for organised knowledge, Frankenstein or Pygmalion? Part 1, *The Electronic Library*, 14 (1), 43-48.
- [4] Clausen, H. (1995) Information quality as a factor in research library internet use, *Proceedings of the 19th International Online Information Meeting*, Learned Information, Oxford, pp. 61-69.
- [5] Clausen, H. (1996) Web information quality as seen from the libraries, *New Library World*(in press).
- [6] *Adgang til og brug af Internettet i Danmark*(Access to and use of the Internet in Denmark). Mediacenter Danmark, May 1996, 30 pp.
- [7] Barkow, T. (Ed.) (1996) Gender and the Web, *Wired*, 4 (4), 70.
- [8] Gvu's 5th WWW User Survey: http://www.cc.gatech.edu/gvu/user_surveys/survey-04-1996/
- [9] URL: <http://www.infokraft.se/si.htm>
- [10] Rupley, S. (1995) Do you fit the Internet profile? *PC Magazine*, 14(20), 30.
- [11] Arnold, S.E. (1996) *Publishing on the Internet: A New Medium for a New Millennium*, Infonortics Ltd., Calne, Wiltshire.
- [12] Pescovitz, D. (1995) The future of telecommuting, *Wired*, 3 (10), 68.
- [13] Bang, T., H. Clausen, S.E. Dyrskov, C.G. Johannsen and O. Johnsen (1996) *ISO 9000 for Libraries and Information Centres: A Guide*. Report of a project supported by NORDINFO. FID, The Hague (FID Occasional Paper 13). Published in Danish in 1995 by NORDINFO.
- [14] Clausen, H. (1994) ISO 9000 and all that: is the information sector ready for the big quantum leap? *Proceedings of the 18th International Online Information Meeting*, Learned Information, Oxford, pp. 37-46.
- [15] Floridi, L. (1996) The Internet: which future for organised knowledge, Frankenstein or Pygmalion? Part 2, *The Electronic Library*, 14 (1), 49-52.
- [16] Kelly, S. and D. Nicholas (1996) Is the business cybrarian a reality? Internet use in business libraries, *Aslib Proceedings*, 48(5), 136-144.
- [17] Costers, L. (1994) The electronic library and its organizational management, *Libri*, 44(4), 317-321.
- [18] Bauwens, M. *Cyberspace, Virtualisation, and the Role of Cybrarians*. URL: <http://www.ecom.be/SPACE.TXT>
- [19] Pescovitz, D. (1995) The future of libraries, *Wired*, 3(12), 68.

Note

- [1] **Serendipity.** A happy coinage by Horace Walpole to denote the faculty of making lucky and unexpected finds by accident. In a letter to Mann (28 January 1754) he says that he formed it on the title of a fairy story, *The Three Princes of Serendip*, because the princes 'were always making discoveries, by accidents and sagacity, of things they were not in quest of.' Serendip is an ancient name for Ceylon (Sri Lanka). The Wordsworth *Dictionary of Phrase & Fable*, 1993.



Missing links — user needs vs. producers' offerings

Sarah Fahy

J.P. Morgan, UK

Introduction

In this session I hope to give an overview of how I perceive providers could better serve the information community. There are a wide variety of reasons which prevent users from making the most of the products on offer: here I will attempt to identify these and then apply them to the variety of sources commonly used within a financial environment. I would stress that all the opinions in this talk are my own, and I am sure each information user in the room would have their own views on this rather controversial topic.

At first sight, it appears that the title to this talk should lead me to provide a long list of databases I would most like to see on the market. All the databases that would make my life easier. However, upon closer inspection it appears that the subject is more complex than this.

The problem

All practitioners know that 80% of user needs can be met if one has unlimited resources, unlimited financial budgets to purchase key research; unlimited time to get the job done. Our problem, and one of the main reasons many of us use online services so heavily, is that we don't have unlimited resources. Users need answers quickly; they need to get value for money now!

So the practitioner invariably must access his or her source rapidly and produce a cost effective result. The issue is accessibility.

How easy is it for me to retrieve the data I need? How cost effective will my results be? How good is the information I am retrieving, and have I accessed the best source? These are the questions that confront the practitioner.

Five key issues

In considering what users need to bridge the missing links, it therefore seems appropriate to look at the key data types most commonly used in a large information unit and consider them with regard to five key issues that determine accessibility:

1. Quantity	is there enough data on the subject? is there too much data?	the producer's responsibility the user's responsibility
2. Quality	how reliable is the data? how sensibly is it presented to avoid confusion?	the producer's responsibility the producer's responsibility
3. Functionality	how easy is it to use the database interface? how consistently and thoroughly is the data indexed?	the producer's responsibility the producer's responsibility
4. Fragmentation	are all the key sources available from one source?	the producer's responsibility
5. Cost	can I afford to access this data? does it represent value for money?	the user's responsibility the user's responsibility

It is the combination of these five factors that will ultimately determine the success of the practitioner — and as we see, very little is actually within the control of the practitioner. The practitioner must keep up to date with the myriad of information sources available and try to keep tabs on changes in coverage. Keeping one's skills up to date is a constant process and has recently become even more vital as more and more products are placed on the end-user's desktop. It is very easy to lose touch with a product if one is not using it day in, day out, but this knowledge is vital as the information practitioner begins to embrace the role of on-site trainer for such

products.

In a large financial institution it is possible to categorise the types of data used as follows:

Press sources	online newspapers, historic and real time such as Reuters Business Briefing, FT Profile, NewsEdge.
Mergers & Acquisitions data	databases giving financial details of deals, joint ventures etc., such as Acquisitions Monthly or SDC.
Financial information	public or private company accounts data such as Disclosure or D&B records.
Market Research reports	industry surveys providing industry structure data, market shares and competitor lists available through the MAID database or individual files on the major online hosts.
Ownership data	shareholder details of companies.
Equity and Bond data	share prices, volumes, market values, bond redemption yields etc.

I have omitted the broad group of specialist sources which cause so many of us problems, such as pharmaceutical or telecommunications sources, as issues vary widely according to subject.

Applying my five 'accessibility' categories to these basic types of data, I would conclude the following:

Press sources

1. Quantity Lots! Too much even, as there is often duplication of stories.
2. Quality Questionable as a source of facts, but essential for market comment.
3. Functionality Here is where the crunch comes! There is so much information available good indexing is essential to aid fast, accurate searching. Too often this is lacking.
4. Fragmentation Becoming a problem as sources like the Financial Times and Wall Street Journal are at present available through limited hosts.
5. Cost Generally users feel they get value for money.

Mergers & Acquisitions data

1. Quantity Limited number of providers, but not a problem as quality is the key issue.
2. Quality Generally poor. There is no consistently accurate source, which means users need access to all available sources, including press articles, to corroborate figures.
3. Functionality Good. The specialist databases exploit the natural format of the data to allow various methods of searching.
4. Fragmentation. Due to the limited number of specialist providers, not a problem. Only an issue when one resorts to press articles.
5. Cost Expensive.

Financial information

1. Quantity Public information — good.
Private company data — sketchy according to country.
2. Quality Public — generally good as it is based on actual accounts.
Private — poor, unsafe. Often only sales figures available.
3. Functionality Public — unless one is accessing the full accounts, the spreadsheet options often offer poor functionality. Spreads may be standardised across countries resulting in misleading figures.
For those providers who do grasp the nettle and supply full financial details, the search interface is often so complex that users who are not accountants need customised reports and spreadsheet.
Private — even where searching for private companies within the same industry, one invariably finds the assigned SIC codes too broad.
4. Fragmentation As more spreadsheet systems become available, one is faced with the problem of standardisation discussed above!
5. Cost Varies.

Online Information 96 Proceedings

Page 126

Market research data

1. Quantity Lots! Especially for consumer topics. Still poor on industrial subjects such as process engineering!
2. Quality Questionable according to reputation of producer. Nevertheless useful for a snapshot of the industry.
3. Functionality Generally good — the structure of the reports lend themselves to relevant indexing. Improvements could still be made on Maid, though. For example, it would be valuable to be able to search across reports for specific tables of data, i.e. market share of yoghurt in Europe, without having to go through tables of contents of numerous reports (as is possible on Markintel).
4. Fragmentation MAID have done a good job bringing together key sources. Many providers appear to be moving to marketing their products themselves now, though, which is leading to an increase in sources available on CD.
5. Cost Good market research data is expensive to collect and consequently expensive to access either in printed reports or online.

Ownership data

1. Quantity Too little. Institutional owners prevail and only the UK and USA markets are well covered.
2. Quality There are very few sources for this information, so users tend to take what they can get. There is room for more products.
3. Functionality Good. Data is structured well to allow detailed searching.
4. Fragmentation Not enough sources out there to be a problem yet.
5. Cost Expensive, reflecting the difficulty providers have in finding this type of data.

Equity and bond data

1. Quantity Whilst the number of providers is still small, the quantity and coverage of companies is good. This is because the ultimate data providers are the relevant stock exchanges.
2. Quality Generally good, though one occasionally finds some peculiarities caused by erratic loading procedures. It is essential that the user understand what exactly he is receiving and the definitions relevant to the data types, e.g. is volume double or single counted by a particular exchange?
3. Functionality Generally good, and improving!
4. Fragmentation Not an issue.
5. Cost Expensive, especially when one wants to have multi access.

It should be clear from this analysis that providers could do a lot to improve accessibility without even starting to tackle the wish list of future databases!

Wish list of future databases

But what would be on that wish list?

- more accessible data on investment and pension funds;
- a database devoted solely to rankings from key journals such as *Fortune* and *Business Week*;
- a database pulling together all key data sources that go into the company profile: financials,
- share prices, broker recommendations and forecasts.

Or to quote one colleague when I asked her what she would like from providers: more discounts!

The indications are that the missing links I have defined will not be addressed in the near future. A recent Frost and Sullivan (Ref 1) report predicted the following growth in revenue by product type across Europe and suggests that we are all in for more of the same — almost 60% of revenue from online services in the year 2000 is expected to be from financial databases.

Year	Financial %	News %	Sci/Tech %	Business %	Biblio %	Other %
1990	60.6	4.8	7.5	18.5	7.1	1.5
1991	61.1	4.8	7.2	18.8	6.7	1.4
1992	61.3	4.9	6.9	19.2	6.1	1.2
1993	61.4	4.9	6.7	19.7	6.1	1.2
1994	61.2	5.0	6.5	20.2	5.9	1.1
1995	61.0	5.1	6.3	20.9	5.6	1.1
1996	60.7	5.2	6.2	21.5	5.4	1.0
1997	60.3	5.3	6.0	22.1	5.2	1.0
1998	60.0	5.5	5.9	22.8	5.0	0.9
1999	59.6	5.6	5.7	23.5	4.7	0.8
2000	59.2	5.8	5.6	24.1	4.5	0.8

Of late providers seem to have been concentrating their investment efforts on producing end-user front-end systems in an attempt to woo the end-user market, so it would seem unlikely that my more basic suggestions will be addressed. In addition, we should not forget that revenues from historical databases account for a very small portion of the whole industry.

A final plea

Nevertheless, I would conclude with a final plea to providers:

- take more responsibility for the quality of sources available online. Concentrate on the key sources over the numerous copycat style products;
- invest in good, consistent indexing of data;
- be more proactive in helping users update their skills. Offer free training and get to know your client well. In my experience the firms that assign one key account manager to coordinate all the interaction between provider and user better match user needs;
- involve clients more in the development of new services. Use beta site tests to get a real idea of what will make the product valuable to the user.

Finally I would remind all providers of a remark made back in 1994 (Ref 2) which is still relevant today: 'it's not what you've got, it's how you access it!'

References

- [1] Frost & Sullivan: European Online Databases Service Markets, December 1995.
- [1] Monica Horten, Survey of Technology in the Office, *Financial Times*, October 1994.



Final product strategies in purchasing and their impact on the information needs of the organisation

David Inbar

Minet Information Technologies and Ben Gurion University of The Negev, Israel

Abstract: Purchasing must support the final products that the organisation sells in order to have a significant contribution to the organisation. Outlined are four innovative purchasing strategies: capitation contracts, product life cycle cost contracts, affordability and supply chain management. All these strategies are focused on the final products of the organisation. They are top-down approaches as they derive their goals and contracts from the organisation's end markets. The strategies have unique information requirements that are very different in quality and characteristics from the needs of traditional purchasing systems. The information requirements are in some cases part of the advantages of the new strategies and in others problems with regards to implementation. Understanding the information requirements of the new purchasing strategies is a must for customers and suppliers alike. Information will be a resource as well as an entry barrier for suppliers. Suppliers that cannot invest in information or sophisticated information systems will not survive in this modern market and disappear. The purchasing strategies also have impact on information centres and libraries. Of particular importance is the emergence of electronic distribution which have profound effects on the supply chain of information.

Keywords: purchasing, capitation, life cycle cost, affordability, design to cost, supply chain, information requirements, user needs

1. Introduction

Purchasing contributes significantly to an organisation and to its business strategy only if its activities and contracts with suppliers are guided by support criteria for the final products that the organisation sells.

Targets such as 'buying cheap' or 'getting the best contracts with suppliers' themselves have no value. Purchasing has to understand the end markets in which the organisation sells its products and obtains its income, to know the contracts and prices in these markets and accordingly to derive its purchasing goals and forms of contracts.

This approach, which we call the final product approach, has led to the development of a number of innovative purchasing strategies. We will address four strategies: capitation contracts, product life cycle cost contracts, affordability and supply chain management.

2. Capitation contracts

Capitation is:

- (1) contract;
- (2) incentive methods;
- (3) working policy with suppliers.

Capitation: providing all of the customers' needs during a certain period in return for a fixed sum which is independent (or partially dependent) on the quantity supplied, when the consumed quantity is not known in advance.

Logic and motive: adjusting expenditure to income, reducing risks. In many health systems the emphasis has moved from increasing income to reducing costs. This has considerably increased the importance of purchasing.

It does not make business sense to pay the suppliers a fee for a product or service when revenues do not move in parallel. The more products I consume the more I will pay, but my income will not rise accordingly. The discrepancy creates risk.

In the capitation agreement there is a reduced risk for the customer. We sell all the risk or part of it to the supplier. In the later case there is risk sharing. There are full and partial capitation contracts. The risks are both cost growth by itself and also cost growth beyond the growth of income.

In capitation the supplier forms very close ties with the organisation and with the organisation's aims and goals. The supplier and organisation are *us*, not *we* and *them*. Suppliers will not try to increase sales by increasing the number of quantities sold or unit prices but will work with the customer to make processes more efficient and to reduce costs. They now have an interest in reducing costs.

2.1. The benefits of costs reduction

- (1) Reducing costs of materials due to quantity discount and being a sole supplier.
- (2) There is no incentive for a supplier to sell unnecessary goods. He has incentives to encourage the customer's goals and to encourage cost effective usage of his products. The supplier is also an advisor. For example, in the case of the medical system the goal is to improve the patient's health.
- (3) A huge saving in transaction and control costs of both parties. Transaction costs are usually valued at less than their real value.
- (4) Reducing the number of suppliers with which the organisation works.
- (5) The supplier will try to become more efficient and to reduce the organisation's costs.
- (6) Economising the marketing costs of the supplier.
- (7) The supplier has more predictability and certainty in the quantity and timing of supply (as opposed to spot orders).

Capitation leads to work procedures being changed, within the organisation and with its suppliers. Labour requirements will change. There will be fewer people working in daily transactions and control, accounting, purchasing (customer) and sales (supplier). The remaining workers will be of a higher standard and will perform more interesting roles.

Capitation will be throughout the whole system, in every supply chain. In the end, no one in the chain will receive any money if the final customer, the consumer, does not pay.

2.2. Problems

- (A) If the consumption is greater than predicted, suppliers face financial risk. Possible solutions:
- (1) contracts that share risk;
 - (2) cost reduction due to a capitation contract also benefits the suppliers and contributes to its stability.
- (B) The supplier must monitor the customer's organisation logistics and consumption processes.

2.3. Requirements for implementing a capitation contract

- Trust between the parties;
- A sophisticated information system;
- Actuarial information. This information allows us better segmentation of data. The segmentation helps us to draw up more clever contracts;
- Supplier having good monitoring systems.

2.4. When implementation would be feasible and worthwhile

- (1) In cases where we want to match expenditure to income, to reduce the risk level, to reduce costs.
- (2) In cases where there is no direct relationship, through the Bill of Materials, between inputs (such as raw materials) and the end product that the client sells, such as medicines versus medical care.
- (3) When the customer has no interest or capability to consume more inputs than what is being supplied due to a capitation contract. Selling more final products will not necessarily mean consuming more inputs.
- (4) If the percentage of the direct cost in supplies is low, the risk of the supplier being capitationally abused will be low.

Examples: insurer with pharmaceuticals, computer services, information services, bank with forms. In the final case capitation contracts will lower transaction costs that in cheap products are a considerable percentage of the product price.

3. Life cycle costing

Life cycle costing is also called TCO — total cost of ownership.

Life cycle costing (for the customer) includes cost of purchasing, operation (labour and other inputs), maintenance, stoppages and taking out of use. The manufacturer also has manufacturing and development costs. The

concept is that when purchasing we should look at life cycle costs and not only at purchasing costs. We can buy an item cheaply and pay in high operations, maintenance and stoppages costs which mean that we did not choose the correct alternative.

All costs will be combined into one parameter — the life cycle cost of the product/system. Very often we have an interaction and trade-off between different costs. For example electrical appliances which are expensive when purchased but economical in energy; expensive vehicles but with cheap maintenance and running costs (petrol). Expensive automated machine which needs little labour to operate as compared to a less expensive machine which needs far more manual work to run. In complicated systems the purchasing costs of an item are only 25–50% of the life cycle cost.

3.1. Problems

The price of the whole system, the LCC, is not known. We only know the purchasing costs. The other costs are not known. They are spread out over a period of years, which may be many years, and spread throughout several departments in the organisation. Every department values costs according to different methods or data. One of the main problems in the model is estimating future costs (in fact, all costs except the purchasing costs).

Several methods

- (1) Receiving data from the manufacturer. He should be more aware of maintenance and operational costs than anyone else and how many stoppages will occur (usually defined for a year). Obtaining data on expensive equipment can be very expensive (millions of dollars).
- (2) Obtaining historical data if we have similar or exact equipment. If the equipment is only similar, or is from a previous generation, then it will also be more of a problem to estimate costs.
New equipment can have better theoretical performance, making it more complicated and less reliable or more sensitive, with larger life cycle costs.

3.2. Further problems with using this model

- Accounting procedures prevent a realistic assessment of LCC;
- Budgetary procedures prevent transferring budgets between departments and spreading over a number of years;
- Every department, including decision makers, looks at its budget and does not want to reach global optimisation.

The next stage after choosing an alternative is putting elements of our study into the contract agreed with the manufacturer. After the manufacturer supplies us with data on his products, we will demand that he take responsibility for any non-fulfilment. The contract must include a penalty should the performance of the machinery deviate from the original data given by the supplier. The penalties will depend upon the nature of the organisation's operations and losses incurred due to the machine stoppages. The contract will, in fact, connect the supplier to the final product of the organisation.

An Israeli company in the defence sector paid millions of dollars to the Pentagon because it presented costs data that were too low.

3.3. Requirements for implementing a product life cycle costing model

Information about the costs is in the later stages of the product (especially in operation and maintenance).

3.4. When implementation would be feasible and worthwhile

- Whenever the purchase of capital equipment is large enough to justify investing in LCC procedures.
- Whenever one expects that the later costs will be a very significant part of the LCC.

4. Affordability

Affordability is a concept which states that the products must be affordable to the customer in addition to its technical and functional requirements. The product's cost constitutes part of the specifications and is a constraint in the design process which is known as design to cost (DTC) and is not a result of the process. The cost becomes a decision variable in the design process.

In a usual design process the product is designed first and priced afterwards. If the product is too expensive it is returned to the designer to try to reduce costs. Along with the pressure of time, this is difficult and expensive. Most product or system costs occur in advanced stages of the life cycle but are determined in the early stages of design.

- In DTC the product's cost goals are derived from the business plan before design is commenced;
- The goals are allocated to the product sub-systems;
- The designers know, from the beginning, the product costs targets and carry out the design accordingly.

They work closely in concurrent engineering in IPT with purchasing and manufacturing people in order to reach these goals;

- If in a certain configuration they do not reach the cost goals, they change design or costs allocations in the system;
- The method is not a miracle cure. It is possible that at the end of the preliminary design process the cost goals will not be met. The designer can say, for example, that he surpassed the cost targets by 30%;
- Even in this particular case we benefit, as opposed to the present situation. The designer can update the initial design to reduce costs, to justify that the product's performance has been changed to reach the cost target. At this stage there is more time to do this. At the same time, marketing will check whether surpassing the cost will affect the business plan.

It is better to be aware as early as possible that the cost targets will not be met, rather than when we are further ahead with the product and have an emotional, marketing and manufacturing commitment.

4.1. When implementation would be feasible and worthwhile

- In new products;
- In existing products that have marketing problems and where one believes that this is due to the product being too expensive vis a vis its performance.

4.2. Problems

DTC design process is longer and more expensive than normal design. One speaks generally of a 20% increase in costs.

Sometimes there is a contradiction between cost effectiveness and affordability requirements. If cost targets are not realistic, the design chosen should be affordable but not cost effective. I would prefer a machine costing \$1m with a \$10m life cycle cost to a \$2m machine with an \$8m life cycle cost, as I now have a budget of only \$1m.

5. Supply chain management

Supply chain management is the management and control process of the flow of products in the organisation, starting from the supply of materials to the organisation's customer and ending up at the supply of finished products to the organisations' customers. The purpose is to maximise the value of all members of the chain, to gain competitive advantages and to carry this out regardless of organisational or global limits.

The supply chain consists of:

- (1) The suppliers that supply raw materials to the manufacturer;
- (2) The manufacturer that converts raw materials into finished products;
- (3) The manufacturer's distribution network of finished products to customers;
- (4) The organisation's customers;
- (5) Outsourcing companies;
- (6) Additional organisations that are involved in the process such as the forwarder, customs broker, agents and distributors.

Increasing competition and globalisation have forced organisations to view their supply chain to and from the organisation and to combine as one network. The chain, if managed properly, will create a competitive advantage through low costs, larger availability of products and improved service. Organisations understand that examining, analysing and establishing an overall policy enables achieving much better results than viewing each system as a separate entity (the problem of partial optimisation).

5.1. Contracts

Working within the supply chain management concept requires contracts between all members of the chain. The contracts will not be simple contracts of products (or services) exchanged for price but will be multiple channel contracts which will stipulate which party holds stock and how it will be kept, guaranteed supply time, stringent quality demands and information requirements. The mutual high dependency of all chain members means having contracts that stipulate the compensation mechanism if one of the members fails.

5.2. Information

Information is of critical importance as the supply chain enables cost reduction and greater efficiency within its member organisations but increases mutual dependency of chain members. Information requirements — amount of information, updating of information, information reliability — are all very high. Organisations belonging to the supply chain exchange sensitive commercial information which, until recently, was kept strictly in the organisation, such as the long-term supply and demand forecast and the reasoning behind it. In more sophisticated supply chains the information is transferred between members by EDI, industry EDI and private EDI of the system.

Online Information 96 Proceedings

members.

5.3. The strategic meanings of supply chain management

In supply chain management each customer expects a lot more from its suppliers than simply receiving products or services for money.

Relations between the customer and supplier are multi-channelled and include subjects such as stock policy, information requirements and the linking of information systems and delivery promises.

The result is setting up entry barriers for suppliers of goods and services. Manufacturing organisations, forwarders and distribution companies must invest in sophisticated information networks or will disappear from the market. Companies with an isolationist and closed policy will suffer from increased and growing logistic costs which will cause inferiority vis à vis their competitors.

As a result of supply chain, policy will not only be determined on an organisational level but also on an inter-organisational level. In most supply chains there is one dominant player who determines or considerably influences policy. This player is mostly, but not always, the dominant customer.

6. Impact on information requirements

The concepts that we outline originate from commercial considerations and not from the information system side. However, if we study what is needed to implement them and the results of this implementation, it becomes apparent that they have a significant impact on the information needs within the organisation.

The purchasing strategies outlined need meaningful information of different quality and characteristics than today's purchasing systems. Every method has its own special requirements. In capitation we need exact and up-to-date actuarial information of consumer history consumption and suitable control systems which will prevent the method being misused. On the other hand, one of the main advantages of capitation is the reduced need for information of the organisation's routine transactions and activities, therefore reducing transaction costs.

The strategies of life cycle costing and affordability require not only purchasing information but also information pertaining to the life cycle costs of the product. Most of the information is of an uncertain nature but an accurate estimation is necessary for choosing the right equipment. A lot of this information is received from suppliers and their contracts which define the quality of information received and the fines for inaccurate information.

Supply chain management requires a large amount of up-to-date and reliable information from parties outside the organisation. The information is also highly secretive and until recently did not leave the organisation. Many companies transfer the information through fast electronic channels which link the organisations as part of the need for up-to-date information. The information element is so important that without the right information it is not possible to implement the methods or the implementation will fail.

Information will be a resource as well as an entry barrier for suppliers. Suppliers that cannot invest in information or sophisticated information systems will be thrown out of the market and fade.

In the forthcoming lecture we will discuss in detail and define the exact information needed for each purchasing strategy, how it can be achieved and the value of this information. This will be done in comparison to the information needs of traditional purchasing methods.

7. Impact on management

The organisation will have to become more intricate. Even though manpower may need to be reduced, workers' quality and performance will have to be increased. Employees will be dealing more with strategies and less with daily routine and transactions.

8. Impact on information centres

We are starting already to see implementations of final product purchasing strategies in the purchasing and selling of information. Knight Ridder (KR), owner of the Dialog system, now offers the use of its Dialog information services on a semi capitation basis. Subscribers can pay KR for information: a global sum which is a little higher than their expenditures on the Dialog system in the previous year. In return the subscriber is entitled to consume about double the amount of information it purchased last year. Selling and purchasing information on a capitation basis is possible because the lion's share of the costs of producing information is fixed, indirect costs. The proportion of direct costs to total costs is very small. Most of the direct costs are in the distribution stage. Due to electronic commerce and the Internet, we expect distribution costs of information to be reduced, making capitation arrangements even more worthwhile.

Information is a prime candidate for electronic commerce and distribution through electronic networks such as the Internet. We expect that electronic distribution will alter fundamentally not only the supply chain in information but the nature of the information product itself and the economics of information. A full description of these changes is outside the scope of this paper. Certain costs elements of producing and distributing infor-

mation will disappear or decrease sharply, such as paper, printing, inventory and physical distribution. The reduced costs and prices will encourage greater supply of information, overwhelming the electronic supply chain.

As more and more information is provided free through the Internet, the pricing of information products will become a complicated issue for producers, providers and purchasers such as information centres and libraries. A possible consequence of this development is the use of fee for performance contracts in the purchase of information. Organisations will still be ready to pay a premium for proprietary information but only if it delivers the results. The information professional will have to master commercial and negotiations skill and have a better knowledge of his organisation in order to negotiate such contracts.

9. Conclusion and recommendations

The new final product purchasing strategies will have a profound effect on the information requirements in organisations, on processes within the organisation and between organisations, and on management and workers' behaviour. These changes and the emergence of electronic commerce and distribution through networks will also have an impact on the information centre and information specialists. In the future the information requirements of processes and contracts should be defined in a much earlier stage than today's by users and information professionals alike. The information centre and information professional should carefully assess and adapt themselves to the new realities.

David Inbar
Managing Director
Minet Information Technologies
4 Ben Yosef
Ramat Gan 52474
Israel
Tel: +972 3 6702208
Fax: + 972 3 6702210



Chemical structure search on the World Wide Web

Wolf-D. Ihlenfeldt

Computer Chemistry Centre, University of Erlangen-Nürnberg, Germany

Abstract: Chemists have been very active in utilising the World Wide Web (WWW) as an information distribution medium and much interesting scientific chemical information is already offered on it. Various classical text-based search engines have made locating information in Webspaces much easier. However, keyword-based searches are often insufficient for chemists interested in structural features of chemical compounds, especially since the naming of chemical structures is far from simple or unique. Chemists have special search requirements and rely on non-textual structure-oriented search methodology. An increasing number of chemical structures can be found in various computer-readable structure exchange formats as MIME attachments to WWW pages on the Internet. If they are catalogued and made searchable with the structure-oriented search methodology that chemists are used to, they can lead to valuable chemical information sources on the Internet which are difficult to locate with other, text-oriented search methods. We have implemented a system for the collection, re-coding, search and context-aware retrieval of chemical structures from the WWW. We assume it is the first instance of a non-textual search engine for WWW data.

Keywords: Web spider, Web search, Web database, chemical structures, MIME

1. Introduction

Chemical information is unique in its contents. While of course research results are described in words, the information the chemist is most interested in is often non-textual. The focus of interest is the chemical structure — either as a schematic entity depicting the connectivity pattern of the atoms, or increasingly as a three-dimensional object. The latter type creates the expected difficulties of displaying the subtleties of a specific molecular conformation by flat and still mostly monochrome images in printed media.

Nearly all chemical structures are systematically extracted from literature and stored in large databases such as CAS Online or Beilstein Online, together with abstracts and references to the original literature. Structure extraction is performed manually by necessity. These databases are searchable by structural criteria (for example full-structure and substructure search) and guide chemists to the original printed articles. Structure-oriented searches are especially important because the official naming schemes for chemical compounds are extremely complicated, error-prone and, even if the name is constructed according to the rules, far from unequivocal. Therefore, the search for a specific compound by name or name fragment is often impossible, or prone to missing many database entries where a different name has been used.

Compared to traditional printed media, the WWW offers some notable advantages to the chemical community (Ref 1). It has therefore been readily accepted as an interesting new information distribution medium, especially for conference proceedings where the problems of copyright, long-term storage and so forth are not so severe. There have even been virtual Internet-based conferences without any face-to-face meetings. Generous use of colour, video and so on is no problem with HTML pages, but the most important feature only possible in electronic media is the attachment of computer-readable structure files as links to hypertext documents (Ref 2). Following these links, chemists can download the original structures into their desktop computers and examine them in detail with helper applications, for example by rotating three-dimensional structures in molecule viewers. A set of MIME types for chemical structure exchange has been proposed to the standards committees (Ref 3). Even before the standard has gained official blessing, the number of chemical papers on the Web which contain attached structure files has exploded recently. Estimations are that at the time of the conference, their number will have surpassed 10,000 — a number which makes indexing mandatory and an interesting problem. Textual indexing services can be provided by the usual WWW search engines, but there are currently no structure-oriented indexing services (Ref 4). Manual compound registration by the authors is certainly not practicable outside the scope of small conferences (Ref 5). Semi-automatically linking compound name occurrences on hand-picked sites to a manually generated structure database is not much more promising (Ref 6). We therefore set out to determine the possibility of creating a fully automatically generated structure-oriented and therefore non-textual database with information from WWW sites (Ref 7).

The information content of the Internet is completely uncontrolled. Therefore the search for specialised infor-

mation types employed only by experts, such as structure data, has the advantage that the scientific content of the hits is generally much higher than in plain text searches. This is especially notable if chemical compounds are considered whose name happens to be known to a broader audience. Textual searches in WWW search engines for scientific information on vitamins, plant protection agents, drugs and so forth are tedious, because the comparatively few scientific information sources are far outnumbered by pages describing health food, cosmetics and similar merchandise or containing political statements. Text in the form of HTML pages is a medium shared between scientists, commerce and interest groups, and text pages are not automatically classified by type of contents in any search engine we know about. As a contrast, we are currently not aware of a single instance of a structure file not related to science in the wider sense, a few abuses as adornments for conference and departmental home pages notwithstanding.

2. Information gathering

The system we have developed consists of four major components. The first is a kind of Web spider, which traverses the Web in an attempt to locate chemical structure files. The second is a sophisticated converter programme, which unifies the various file formats into a common representation suitable for inclusion in a structure database. A third component, the verifier, checks whether the sources of the database records are still accessible at their original locations and that their contents are unchanged. The final major system component is the structure database proper, which is accessible via Internet channels either by a custom client or by means of a form-based Web interface.

It is fruitless to attempt another complete indexing of Webspace (the sum of all WWW-accessible data). Rather than search randomly, our spider starts around a set of promising start pages and expands them two or more link steps. Promising start pages in this context are pages which contain words indicating chemical content. We currently use a manually generated word list of some 100 terms. Typical words contained in the list are 'chemistry', 'molecule' or 'phenyl'. A start site generation script presents these words to standard textual Internet search engines and extracts the URLs from the replies. Currently the script decodes the answer pages of Lycos and Alta Vista (Ref 8). Unfortunately, none of the bigger search engines allows the complete retrieval of large answer sets any longer. Therefore it is impossible to obtain more than only a small portion of, for example, the URLs containing the word 'chemistry', which is present in more than 400,000 instances in the June 1996 Alta Vista index. By expanding the vocabulary to other more specialised chemistry-related terms and automatically merging the search engine results, we routinely update a database of starting URLs which currently contains about 15,000 entry points (Figure 1).

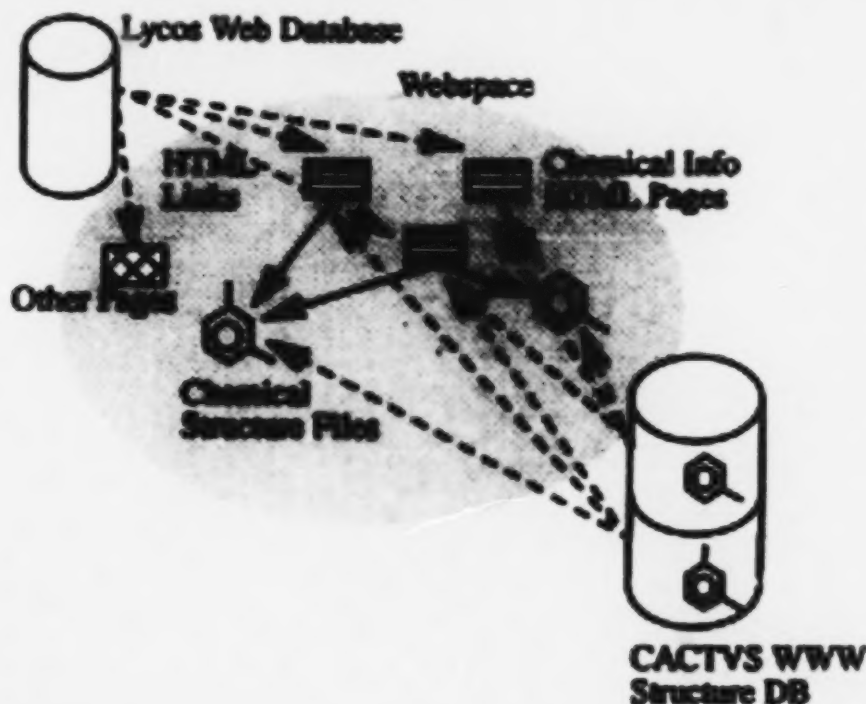


Figure 1: Relationship between textual Web databases (for example, Lycos) and the Web molecule database.

The spider program expands the links in and around these pages up to a specified depth, which is initially set to two for unclassified sites. We do not store the page contents. Only the extracted links and the link texts, plus some administrative information such as the time of the last visit, are collected in a set of gdbm databases. The 15,000 starting points expand to about 750,000 different URLs. Host name aliases are resolved to primary names, but it is technically impossible to detect aliased directories. The gatherer is regularly running at European night-time on a small cluster of three to five workstations, and comprises about 30 independent processes. This is the optimum in the current implementation because after that the read/write synchronisation and locking of the gdbm files becomes the limiting factor. For performance reasons, it is mandatory to cache information from the network wherever possible. This is especially true for host name resolutions, which can take a comparatively long time. If a host name in a URL is in the persistent shared cache and the entry is reasonably recent, the WWW page retrieval is conducted directly with the host's known IP address instead of its name. Another example of useful caching is the standard Web robot exclusion file *robots.txt* from the server root directory. It is cached in pre-parsed form (Ref 9). The spider program, including the client/server communication routines plus all additional scripts used in the information collection phase, were written for convenience in Tcl/TclX (Ref 10). Performance is limited by the network connections and NFS file locking, not the page data analysis.

From time to time, the links to structure files are sifted out from the link database. The basic recognition criteria are the file name suffixes which are more or less consistent and characteristic for the various popular chemical structure exchange file formats. Additionally, a clustering step is inserted here. Sites which contain a notable number of structure files whose URLs have a common part are detected. Another expansion step with an extended link depth of three to five levels is performed, starting with the common directory part. During these scans auxiliary checks are performed to avoid wandering off the cluster site. This second sweep of cluster sites typically yields another set of structures comparable in size to the initially located set. Files suspected to be structure files are downloaded and stored into another gdbm database with their URL as key. At the time of writing, a total of about 3000 structures have been found. This is very little compared to the more than 20 million entries in databases like CAS Online, and the 30 million WWW pages in the Alta Vista index of June 1996. Also, the scanning effectiveness is still small, comparing the number of structure files to the locally expanded 750,000 links in the spider database (0.4%). However, the number of structure files is explosively growing and has been approximately doubling every two months in the last year. We think it is necessary to develop techniques to index this kind of information now in order to be prepared for the arrival of the expected flood of new data of this kind in the near future should the growth continue — and there are no reasons why it should not, especially since now a number of refereed electronic chemical journals have gone online. Three thousand compounds is a number which definitely is no longer manageable manually but small enough to be handled by a simply structured experimental database.

It has been asked why we did not try to locate the structure files directly by textual search with their characteristic endings as query. The main reason is that links to these files are necessarily unidirectional. None of the common chemical structure file formats has provisions to include a URL as a back link to a referring page. The structure itself, while it is the primary vehicle for the location of information for chemists, is of value only in the context of the pages describing its properties and the experiments performed with it. Therefore, not only the structures must be registered but also the links leading to it. These can be located only by a search starting at text pages. Only very recently search engines such as Alta Vista have implemented search capabilities to locate referring pages. Once this functionality becomes stable and the search engines allow (again) the retrieval of an unlimited number of hits, it may be possible to construct the link tree by such queries instead of building it oneself. However, facing a constant overload, these search engines become continuously less responsive and therefore it is not clear whether an approach based exclusively on search engine results would actually be faster than a restricted local Web traversal.

3. Information re-coding

The second phase in the production of the database is the re-coding and unification of the structure files. This is a prerequisite for structure-oriented searchability. This task turned out to be magnitudes more difficult than the collection of the primary data and link relationships. There are about two dozen important structure file types in the MIME set and we have implemented I/O modules for most of them. Unfortunately, many of them — and unfortunately the most popular ones — are not very well suited for the coding of connectivity-based structure representations which are the base for all connectivity-oriented structure search algorithms. Among the structure file formats, the two most common ones are the formats of the Brookhaven Protein Database (PDB) and the XYZ format. The XYZ format is truly minimalist in its information content. It contains only three-dimensional atom coordinates, which are normally sufficient for simple 3D visualisation tasks, but the files do not contain any connectivity at all. PDB files can contain connectivity but often do not, and worse, there are no provisions to encode bond order — a very important concept for structure-oriented searches. We encountered numerous cases where PDB files contained connectivity only for selected parts of a molecule which were related to a certain problem. Furthermore, due to the ill-defined nature of the PDB format which is several decades old and was originally designed for a specific database of proteins, it has often been severely abused to encode data the format was not designed to hold. In PDB files, it is not even possible to determine unequivocally whether an atom labelled CE is an e-carbon atom (of the amino acid lysine, for example) or cerium — the two very different meanings are normally discernible for chemists looking at the overall nature of the compound

but difficult to recognise programmatically in unknown files without context information. Both meanings were actually encountered.

In contrast to connectivity reconstruction, as it is for example performed when electron density patterns from x-ray crystal structure analysis are decoded, several important pieces of auxiliary information are missing in many of the structure records. For example, often the structures are not guaranteed to be electroneutral and the status of hydrogen atoms (whether they are all included, or only at selected atoms, or — following the custom with certain compound classes — not at all) is generally unclear. It was a major effort (about 20,000 lines of code) to implement the machinery which now for more than 95% of the samples reconstructs a structure with a complete hydrogen set connectivity including bond order, stereochemical descriptors and 2D display coordinates which is identical to structures reconstructed manually by an experienced chemist. The system relies in its functionality on the chemical data handling capabilities of the CACTVS system (Ref 11). It uses a multi-step approach. At several processing stages the introduction of heuristics was unavoidable. They were derived from general chemical knowledge, geared to the problem domain. The most important processing steps are, in the order listed, atom type recognition, hydrogen status guess, generation of the first set of single bonds, ring detection, tentative assignment of aromaticity to rings, aromatic ring Kekulé structure generation, multiple bond order detection, radical and charge equilibration, neutralisation of improbable charges and radical centres, display coordinate generation, stereo centre detection, stereo descriptor assignment, wedge bond assignment, and finally the generation of display attributes for double bonds inside rings, suppressed hydrogen atoms, carbon atoms displayed by convention only as node and not as symbol etc.

Reconstruction of structures is currently a batch process. The time scale ranges from less than a second for small molecules without complex aromatic ring patterns to half a minute for large proteins (several thousand atoms) with very incomplete information. These figures were measured on a 200 MHz SGI Indigo II workstation. All structures are converted into a common format and augmented with display coordinates, stereochemical descriptors, names (if present in the original file), the original URL and the two-level tree of referring URLs. We do not retain the original file after conversion. We also compute a group of very discriminative 64-bit hash codes for all structures (Ref 12).

The hash codes are used by the verifier and for full-structure searches. The verifier is a separate program which regularly checks whether the structures in the database are still accessible and have not changed. The verifier operates first by downloading the structures using the stored original URL. If successful, it repeats the conversion process and compares the hash codes of the original and the database entry. In case of changes, all URLs in the link database which share a common name part with any of the URLs referring directly or indirectly to the changed or disappeared structure are marked for a rescan. This strategy helps to detect new structures because often the contents of changed sites were globally edited, not just with respect to the single structure where the change was detected.

4. Database access

The database is managed by a customised Postgres95 object-relational database server (Ref 13). It employs dynamically loaded custom function modules to provide the chemistry-specific structure search facilities. The implementation of the substructure search follows the traditional model. It uses a 256 bit fragment-based screening vector which filters out most of the records in the database from the candidate list. Only the survivors are submitted to an atom-by-atom matching step (Ref 14). Full-structure search is provided without a subsequent atom-by-atom match by simple comparison of one member of the aforementioned set of 64-bit hashcodes. The hashcoding procedure is so reliable that there are currently no known examples of collisions. The algorithm has been subjected to tests with documented problematic cases from the literature (Ref 12). The hashcode is pre-computed for the database records in several variations, such as with inclusion or omission of stereochemistry, which correspond to the various full-structure search modes.

The performance of the database, though certainly not optimised for this kind of application, is sufficient for the current number of entries. Typical retrieval times are from one to five seconds on an Indigo II workstation. A noteworthy feature of our database is that there are no molecule size limitations. The molecule sizes in this dataset vary dramatically, from two to several thousand atoms. Commercial molecule database products typically impose a rather low limit on the number of atoms they can handle, and many structures from our comprehensive collection could not be stored in such a database. In our approach, the storage model of the structures in the database is dynamic and transparently switches from an opaque field in the structure table proper to an external large object if a size threshold is exceeded.

The structure database is freely accessible via the Internet as a research prototype. We provide two basic methods of access: by means of a custom client, and from a standard WWW page.

The custom client is part of the CACTVS tool distribution and can be obtained free of charge for the major Unix operating systems (Ref 15). The advantage of the client is that it is tightly integrated with other tools from the distribution and can use complex means of graphical visualisation. Drag and drop of structures and search fragments in graphical representation and other convenient operations are supported (Figure 2). This degree of functionality is impossible to achieve with standard HTML pages. From within the client, the Mosaic WWW browser can be instructed via its CCI interface to display the referring text pages (Ref 16).

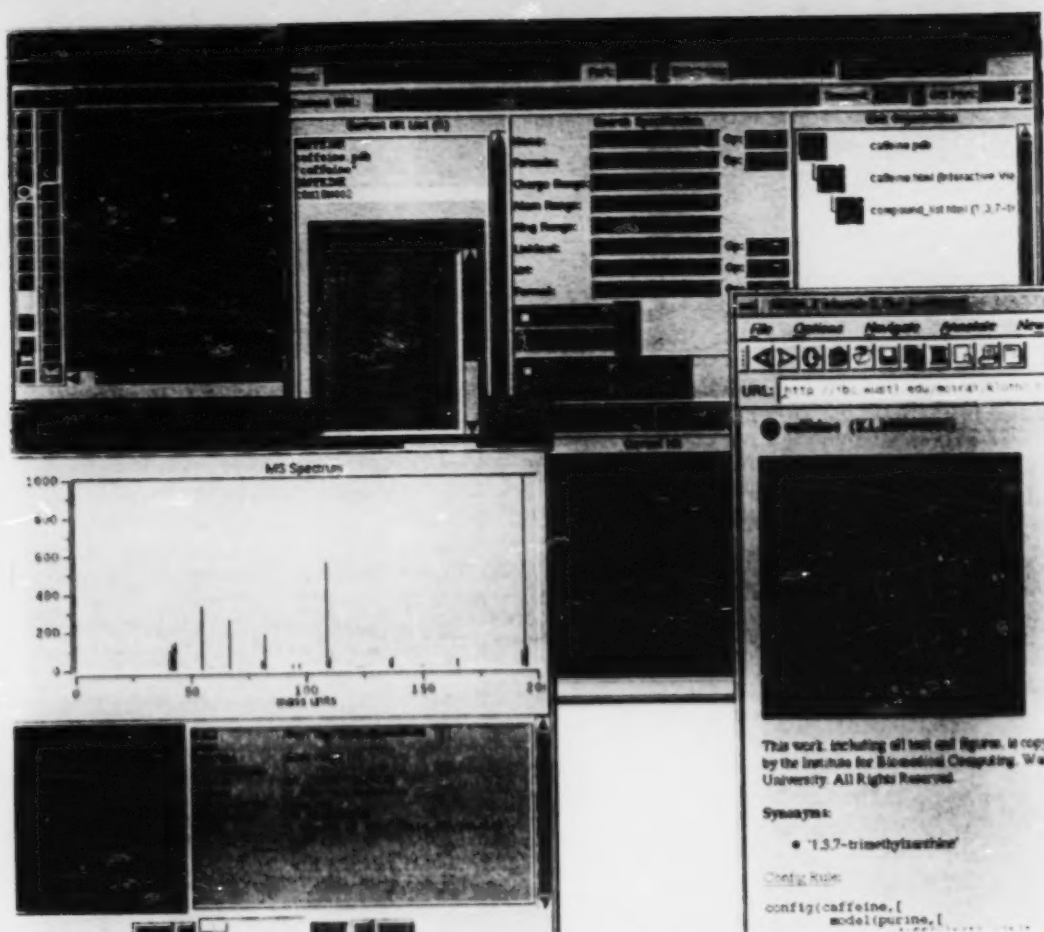


Figure 2: Searching for chemical information about caffeine on the WWW with the database client (upper right). The structure of caffeine was drawn in a structure editor (upper left) and dragged into the full-structure search pocket of the client. Five hits were found. By selecting one of those hits (highlighted yellow), the corresponding structure is retrieved and its link context displayed in the right-hand window of the client. Referring HTML pages can be displayed in a standard WWW browser by selecting them from the context tree (lower right, on this page the biochemistry of caffeine is elucidated). Chemistry-specific information is viewed with helper applications. As an example the mass spectrum of caffeine is displayed on the lower left.

Access by means of the alternative forms-based WWW interface (<http://www.schiele.organik.uni-erlangen.de/services/webmol.html>) is somewhat less convenient, but avoids the use of specialised software (Ref 17). Search fragments must be input manually as SMILES strings, or drawn with an external structure editor and pasted into text fields as ASCII structure files. Response pages are dynamically generated by CGI scripts on the server side. These pages include GIF images of the structures matching the query, so this mode of access generates a notably bigger load on the network than the compact transfer of the information to the custom client, which handles structure display and query command status updates autonomously. The CGI program at the server side is essentially a GUI-less version of the client program. It outputs HTML pages and GIF images instead of the normal GUI-based query result representations. Both programs are scripts written in Tcl/TclX and are executed on a general-purpose interpreter program for chemical structure handling. The only major difference between the two interpreter programs is that the client version is linked with the Tk GUI toolkit and the CGI program is not (Figure 3).

Structure Search Detail

Name	TSJB
Formula	C ₅ H ₇ BO ₂
SMILES	C=COBO=C
Weight	85.071 g/mol
Atoms	13
Rings	0
Charge	0
Orig. File Format	Pdb
URL	http://www.ch.cam.ac.uk/3D/MM
References	1

Structure:

Retrieve from Database: Format: XYZ

Link Environment

- boatB.pdb
- aldol2.html ()
- aldol1.html()

The Boron Mediated Aldol Reaction

Introduction

These pages describe an investigation of the boron-mediated aldol reaction. It was undertaken by Dr. John E. McGrath and Dr. Andrew G. Davies at the University of Cambridge, and by Professor Cesare Geronzi and Professor Anna Bernardi at the University of Milan. All the work described here has been published.

The boron-mediated aldol is a useful reaction in organic synthesis, because it can form two new chiral centres whilst forming a carbon-carbon bond.

The absolute sense of the induction can be controlled by chiral ligands on boron, by a chiral auxiliary or by a chiral catalyst. (C. Geronzi, A. Bernardi, J. E. McGrath, A. G. Davies, *J. Chem. Soc. Chem. Commun.* 1996, 1191)

Figure 3: Searching for information in WWW documents about the mechanism of boron-mediated reactions. Here the forms-based WWW interface is employed. Details about a specific hit are displayed in a dynamically generated HTML page (left). The link environment of the hit is shown on the lower left. The icons are linked to the original pages and an original paper was loaded into another browser window on the lower right. The structure can be retrieved as 3D model for close inspection with the aid of molecular visualisation helper applications (upper right).

5. Query Example

A search for caffeine or caffeine yielded more than 10,000 matches in the Alta Vista June 1996 catalogue (Figure 1). As expected, truly chemical information is extremely sparse among these hits. The corresponding structure-based search in our catalogue resulted in only five hits. A string search on the URL of the registered files yielded another hit: The mass spectrum of caffeine was coded in a JCAMP-DX file without an attached structure. All references contain information highly relevant for chemists and therefore a complete suppression of irrelevant information was achieved. However, an examination of some of the hits obtained by the keyword search revealed that not all interesting information had been retrieved, since there are a number of information sources dealing with caffeine and its chemistry which include drawings of formulae and reaction schemes as images only, and do not yet link structure files to their texts which could be indexed by our robot. It remains to be hoped that once the value of structure attachments for chemistry-based indexing becomes more widely known, the majority of sites focusing on chemists as clients will provide these attachments. Since non-scientific information on the WWW is certainly growing more rapidly than scientific information, the selectivity of text-based search engines will continue to degrade. Automatically indexed structure attachments are a convenient method for information providers to help them to stand out and to draw visits from the targeted scientific readership.

Online Information 96 Proceedings

6. Summary

We have implemented a structure-oriented search and retrieval system for chemical compounds found as MIME attachments to textual HTML information on the WWW. As far as we know, this system constitutes the first non-textual search engine for WWW information (Ref 18). We expect that the importance of non-textual specialised information will grow very rapidly for the Internet community. We have demonstrated that the collection, re-coding and database production of molecular information from the Internet is not trivial but absolutely feasible. Searches on this data provide valuable information for users with scientific interests which is not available, or much more on target than results obtained, by other available Internet search methodology.

Dr Wolf-Dietrich Ihlenfeldt
Computer Chemistry Centre
University of Erlangen-Nürnberg
Nägelsbachstraße 25
D-91052 Erlangen
Germany
Tel. +49 9131 856579
Fax +49 9131 856566
E-mail: wdi@eros.ccc.uni-erlangen.de

Acknowledgements:

This work was supported by FIZ Chemie, Berlin.

References

- [1] Rzepa, H.S., B.J. Whitaker and M.J. Winter (1994) Chemical applications of the World-Wide-Web system, *Journal of the Chemical Society, Chemical Communications* 1907-1909.
- [2] Krieger, J.H. (1995) Chemistry sites proliferate on the Internet's World Wide Web, *Chemical and Engineering News*, **73**(46), 35-46
- [3] Chemical MIME proposal: <http://www.ch.ic.ac.uk/chemime2.html>
- [4] Ihlenfeldt, W.D. and J. Gasteiger (1995) Beyond the hyperactive molecule: search, salvage and visualization of chemical information from the Internet, in L. Hunter and T.E. Klein (Eds.) *Pacific Symposium on Biocomputing* 96, World Scientific, pp. 384-395.
- [5] The ECTOC Hyperglossary: <http://www.ch.ic.ac.uk/ectoc/glossary/>
- [6] The ChemFinder database: <http://chemfinder.camsoft.com/>
- [7] Ihlenfeldt, W.D. and J. Gasteiger (1996) Treasure hunt: on the track of reusable structural information in the World Wide Web, *Software Development in Chemistry* 10, Springer, in print.
- [8] The URLs of the cited search engines are:
Alta Vista: <http://altavista.digital.com/>
Lycos: <http://www.lycons.com/>
- [9] Robot exclusion standard: <http://info.webcrawler.com/mak/projects/norobots.html>
- [10] Ousterhout, J. (1994) *Tcl and the Tk Toolkit*, Addison Wesley. See also <http://www.sunlabs.com/research/tcl/docs.html>
- [11] Ihlenfeldt, W.D., Y. Takahashi and H. Abe (1994) Computation and management of chemical properties in CACTVS: an extensible networked approach toward modularity and compatibility, *Journal of Chemical Information and Computer Science*, **34**(1), 109-116.
- [12] Ihlenfeldt, W.D. and J. Gasteiger (1994), Hash codes for the identification and classification of molecular structure elements, *Journal of Computational Chemistry*, **15**(5), 793-813.
- [13] Postgres 95 database home page: <http://epoch.cs.berkeley.edu:8000/postgres95/index.html>
- [14] Hicks, M.G. and C.J. Jochum (1990) Substructure search systems, 1. performance comparisons of the MACCS, DARC, HTSS, CAS Registry MVSSS and S4 substructure search systems, *Journal of Chemical Information and Computer Science*, **30**(2), 191-199.
- [15] CACTVS system home page: <http://www.schiele.organik.uni-erlangen/cactvs/>
- [16] Mosaic CCI interface: <http://www.ncsa.uiuc.edu/SDG/Software/XMosaic/CCI/cci-spec.html>
- [17] WWW chemical structures database Web page: <http://www.schiele.organik.uni-erlangen.de/services/webmol.html>
- [18] Other interesting non-textual information sources on the WWW are maps, VRML scenes and images. See for example Wilensky, R. (1996) Toward work-centered digital information services, *Computer*, **29**(3), 37-44.



Chemical publishing on the Internet: electronic journals — who needs them?

C.J. Hildyard and B.J. Whitaker

School of Chemistry, University of Leeds, UK

Abstract: The rapid growth in the use of the Internet in recent years has in large part been due to the interest in and use of the World Wide Web. Indeed, the expansion in this form of electronic communication is so striking that it has led some commentators to make comparisons between the development of the Web and the introduction of the printing press in fifteenth century Europe. As a medium for publication and information dissemination, superficially at least, electronics promises much. Using mechanisms built into the structure of the Web, such as hypertext, information sources can be linked across the globe. So called spiders and robots can then 'crawl' the Web to construct indexes and other search tools. Thus a reader can in principle locate and access information related to a specific query from servers located anywhere in the World. The ability to embed multimedia files into Web documents (in either HTML or HTF) also offers authors techniques with which they can extend what can be said in words alone. In this paper we examine the possibilities offered by the World Wide Web and related technologies in the area of scientific publication. Work within the Electronic Libraries Program (eLib) project to establish an electronic version of the Royal Society of Chemistry's journal, *Chemical Communications*, is described. We focus in particular on issues related to document preparation from SGML master documents, and the development of multimedia methods for the presentation of complex scientific data, such as three-dimensional molecular structural information, using MIME and Java. The value of the Web for scientific and technical publication is clear, but the structure of the Web and current modes of use also raise difficult issues, and in the conclusion we attempt to assess critically the impact of electronics on the publishing process itself.

Keywords: electronic publishing, chemistry, World Wide Web, markup languages

1. Introduction

There are many issues to consider when developing an electronic version of a journal, from the perspective of the publisher, the author and the reader alike. The opportunities for all are obvious. From the publishers' point of view, to name but a few, there are opportunities to speed up the process of publication, to create 'value added' products using multimedia and to monitor document access and usage by readers. From the point of view of the author and the reader the attraction lies in the novel ways in which data can be presented and interacted with. More fundamentally, the World Wide Web (WWW) enables access to a wide variety of information on-demand, conveniently and efficiently.

The drawbacks of the electronic medium can in the main be put down to lack of precedent. The technology is relatively new and is developing so rapidly that there has simply not been time to legislate for problems such as copyright, privacy and data protection on the Internet. However, the obvious promise that this medium offers has led to a number of projects experimenting with the possibilities of electronic publication. It is to be hoped that between them, a framework for electronic publication can be worked out which addresses these issues without being unduly restrictive. The following is a brief description of the aims, strategy and experience of one such project.

2. The CLIC consortium

The CLIC consortium is a project within the Electronic Libraries (eLib) programme which aims to establish an electronic version of the Royal Society of Chemistry's flagship journal, *Chemical Communications*, to run in parallel with the printed version (Ref 1). The eLib programme itself was set up by the Joint Information Systems Committee (JISC) following the recommendations of the Follett report (Ref 2) into the use of information technology (IT) in the electronic library. A key motivation is to investigate whether IT can help to alleviate some of the problems of university libraries today—principally the ever spiralling cost of journal subscriptions and the

associated storage costs. Although mainly used by structural and organic chemists *Chem. Comm.* is a general chemistry journal which aims to disseminate rapidly internationally important research results. The journal takes great pride in its speed of publication, averaging 105 days from receipt of manuscript to publication in 1995, and in the high quality of its content and refereeing (rejection rate ca. 50%). The journal is published biweekly and each issue typically runs to over 150 printed A4 pages containing some 80 articles. Each issue also contains a longer feature article highlighting the work of a leading chemist.

Whilst developing a strategy for the implementation of an electronic publication, several issues have arisen. Three important areas will be discussed here.

- (1) document archiving — the electronic storage of the research papers;
- (2) document delivery — the method by which documents are provided to the user;
- (3) 'added value', exploiting to the greatest possible extent the features of the WWW, providing information which by its nature could not be conveyed by the printed page.

2.1. Document archival

Papers are submitted to the Royal Society of Chemistry (RSC) publications offices in Cambridge on paper and in electronic formats, such as Microsoft Word, WordPerfect and to a lesser extent TeX and LaTeX. The RSC encourages authors to submit manuscripts on disk or by e-mail, but currently only about 10% of submissions are so received (although about 40% of submissions are accompanied by discs containing graphical information such as figures, etc.). Of these the vast majority (>90%) are in Microsoft Word format. The rapid development of document preparation software, however, ensures that at any given time there are dozens of formats concurrently in use by potential authors. Obviously the transfer of information between computer systems relies on the compatibility of these formats (which can cause headaches for the publisher). More crucially the constant development of document formats posits a fundamental difficulty for the long term archiving of electronic documents due to the possibility, not to say probability, of formats becoming obsolete. A publisher who is naturally responsible for the long term storage as well as the dissemination of such electronic data therefore faces a real problem.

The solution adopted by the RSC, along with many other publishing houses, has been to make use of Standard Generalised Markup Language (SGML) (Refs 3, 4). Markup is the formatting information which tells a word processor or display client how the document should be rendered, for example font type and size, alignment, etc. Sections of text are usually marked up by placing the contents between two delimiters or tags. Users of TeX (Ref 5), in which the markup is inserted manually, will be more familiar with this concept than those who use WYSIWYG software, in which the markup is largely invisible. SGML is an alternative type of markup that is descriptive rather than procedural. It defines *containers* which specify the semantic content of the textual elements rather than their appearance. Each SGML document starts with a *Document Type Definition* (DTD), which lays out the rules for the markup of a document in terms of the elements it contains. This allows the markup to be designed with the specific data types likely to be encountered in mind — a feature which is particularly useful in specialised subject disciplines such as chemistry and one to which we shall return later.

To give an example of descriptive vs. procedural markup, in SGML this paper might be marked up as:

```
<!DOCTYPE article PUBLIC "-//Online/DTD article//EN">
<ARTICLE>
<TITLE>Chemical Publishing on the Internet ... </TITLE>
<ADDRESS>School of Chemistry, University of Leeds ... </ADDRESS>
<AUTHOR>C.J. Hildyard and B.J. Whitaker</AUTHOR>
<ABSTRACT>The rapid growth in the use of the Internet ... </ABSTRACT>
<SECTION>
<SECTIONHEADING>Introduction</SECTIONHEADING>
There are many issues to consider when developing ...
</SECTION>
...
</ARTICLE>
```

as opposed to, say

```
{\rtf1\ansi\deflang1033{\fonttbl{\f4\froman\fcharset0\prq2 Times
New Roman;}{\f8\froman\fcharset0\prq2 Times;}}
{\stylesheet{\f4\lang2057 \snext0 Normal;} ...
\l4\lang2057 {\b\fs28 Chemical Publishing ... \par }
\par \pard C.J. Hildyard and B.J. Whitaker
\par
\par School of Chemistry, University of Leeds, Leeds LS2 9JT, U.K.
\par
\par \pard \qj {\b Abstract}
\par \pard \qj \tab The rapid growth in the use of the Internet ...
\par \pard \qj
\par \pard \qj {\b 1.\tab Introduction}
\par \pard \tab There are many issues to consider when developing ...
```

which is the way the document appears marked up in Rich Text Format (RTF). The latter contains information about the font size, character set, etc. and defines the way in which the document should appear on the page in addition to its content.

In SGML, on the other hand, the markup simply defines the type of each textual element. The first line contains a *document type declaration* identifying the DTD required for the document — in this case a (fictitious) declaration for a document in the style of an English language (EN) article for these proceedings. This is followed by a start tag <ARTICLE>, identical to the tag used for the DTD declaration, which is called the base document element and which activates the DTD. The elements following the <ARTICLE> start tag are fairly self-explanatory. Finally the end of the document is identified with a closing end tag </ARTICLE>. Notice that only the logical structure of the document is important, for example the <ADDRESS>...</ADDRESS> container appears out of sequence from that in the printed version of this article. Textual elements may be nested so, for example, we could define a container <CITATION>...</CITATION> with sub elements <Author>...</Author>, <Year>...</Year>, <Pages>...</Pages> and so on. SGML tags may be also qualified by <|attributes as in <CITATION type=Book> which might make certain sub elements mandatory, e.g. <Publisher>. These might be optional for other attributes, e.g. with <CITATION type=Journal> it would be unusual to specify the publisher. Each SGML document, known as an SGML *document entity*, consists of three components: an SGML declaration defining the syntax and character set used within the document, the DTD defining the logical structure of the document, and the text itself.

Since the markup in SGML defines the logical structure of the document rather than its appearance, the danger of the data becoming obsolete is significantly reduced — the philosophy being that whether a document is rendered on paper or on a CRT its basic structure is immutable. The final appearance of the document is determined by a post-processor. In our work this is done using CoST (Ref 6). CoST is a Tcl-based SGML post-processing tool, built on sgmls, a public domain SGML parser developed by James Clark (Ref 7). Restricting the formatting stage to a post-processor rather than including the information along with the data itself has clear advantages. Specifically in the context of publishing on the WWW it addresses the problem caused by the fact that the 'standard' Web markup, Hyper Text Markup Language (HTML) is anything but a standard. Currently at version 2 HTML (Ref 8) lacks significant functionality when it comes to rendering scientific and non-English language text. For example, the lack of support for Greek characters creates difficulties in properly rendering scientific documents, particularly mathematical formulae.

The solution adopted by most publishers, including ourselves, is to link to a library of GIF images of the characters in question. This is a crude kludge. It works but it has unfortunate side effects. It adds to the overhead on document transfer and significantly affects download times adversely. Furthermore, since font size is both browser and context dependent (e.g. headings are rendered in a larger font than normal text), there is no way of knowing whether the image supplied is the appropriate size. This can lead to uneven line spacing and a poor overall appearance on the client, and from our limited 'market testing' to date is unpopular with readers.

The latest implementation of HTML, HTML3.2 (Ref 9) (at the time of writing still a proposal) does not address the problem of character sets either, although some extra functionality is added. The question of standards is an important one. HTML 3.2 is not a standard as such but a specification drawn up by the World Wide Web Consortium (W3C), with input from various vendors including Netscape Communications Corporation, Microsoft, IBM, Sun Microsystems, Novell, SoftQuad and Spyglass. The likelihood is that standards will be replaced by periodically updated specifications of the current practice in browser technology. Developments are therefore sure to be rapid and publishers must be able to react swiftly. This is an unsatisfactory situation and certainly not one that a major commercial publishing house can easily accept. By maintaining the document in SGML form and subsequently converting it to HTML, changes in the definition and functionality of HTML can be handled by altering a few lines of code in the translator rather than the source data itself. This not only protects the integrity of the database in a recognised ISO format but allows a publisher to react quickly to developments in WWW servers and browsers. The document archive is in a sense 'future proof'.

The DTD used in the CLIC project is a modified form of the ISO standard 12083 DTD which was designed to cater specifically for scientific and technological applications (Ref 10). However, in order to handle the vast array of chemical information and data types, the DTD needs to be extended. Chemistry Markup Language (CML) (Ref 11) is a proposal designed to address this problem by defining semantically rich chemistry-specific markup to represent the diversity of chemical data as completely as possible. It is only with a sufficiently 'tag-rich' DTD that one can properly make use of the variety of possibilities offered by the Web. We expand on this point in Section 2.3.

2.2. Document delivery

Once the SGML archive has been established one must decide on the method of delivering the document to the user. It would be counter-productive to have an SGML database and an HTML database containing the same information in existence at the same time. This also goes against underlying principle of any database, namely that there should only be one master. The approach used in the CLIC project is to generate the HTML 'on-the-fly' using CoST via a *Common Gateway Interface* or CGI script. This is executed as and when the user requests an article, and removes the necessity for a parallel HTML archive. Although conversion times are rapid — 1-2 CPU seconds on a Sparc2 for a typical *Chem. Comm.* paper — the converted articles may be temporarily cached, so that a future reader requesting the same article a short time later does not have to wait for the conversion process to be repeated. If the lifetime of documents held in the cache is sufficiently short, updates to the code will filter through to the user very quickly. Thus there will be only one permanent archive and a dynamic

cache of temporary documents, in the interests of efficiency and speed of delivery.

CoST operates on the Element Structure Information Set (ESIS) which is produced by sgmls and not the SGML document instance itself. This ensures that the SGML is validated against the DTD before being processed, providing an extra safeguard for the integrity of the archived SGML data. SGML documents can be seen in terms of a tree structure which CoST navigates using a powerful query language. The contents of each tag in the SGML document instance are associated with a corresponding HTML tag, with hyperlinks generated where necessary. As long as there is sufficient information within the SGML document instance, or in other words the DTD allows for a comprehensive markup of the document structure, any HTML features or indeed non-HTML external applications can be generated automatically. CoST can also be used to generate the Graphical Abstracts pages which will act as the 'front door' for the electronic version of *Chem. Comm.*, in line with the structure of its printed equivalent. Figure 1 illustrates the overall structure of the conversion process. Using this model, one can develop further tools to convert the archive to other formats, as and when they are needed. This allows the publisher to provide a variety of different versions of the same information, according to the taste of the user, without having to store huge amounts of duplicated data.

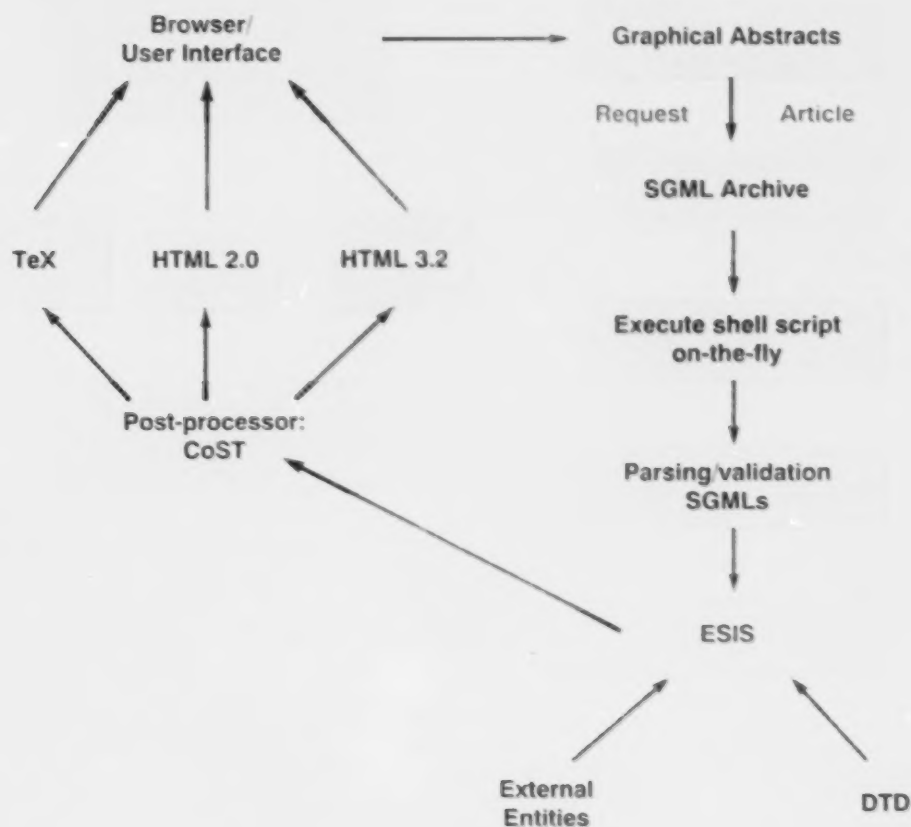


Figure 1: Outline of the document delivery process. The user accesses the Graphical Abstracts page, requests a full article and receives a formatted version via on-the-fly conversion from the SGML archive using the Tcl-based post-processing tool, CoST.

2.3. Added value

An electronic version of an article which is identical in both content and layout to the printed version is an opportunity missed. From the publisher's point of view it is both uneconomical and largely pointless to produce two

parallel versions of the same thing. The WWW offers exciting opportunities, in particular in the field of chemistry, to present data in ways which are impossible on the printed page: to 'add value' to the article concerned. Examples of added value could be to embed interactive 3-D molecular viewers, using Virtual Reality Markup Language (VRML) or to use a mechanism based on *Multi-purpose Internet Mail Extensions*, or MIME, types to launch external applications automatically when a piece of embedded data is accessed by the user. NMR spectral data, reaction schemes and equations can also be treated similarly. We have called this concept 'hyper-active molecules' and demonstrated it in a number of applications (Refs 12-14). Indeed, one can envisage any object in an article becoming an interactive feature, e.g. equations could be dynamically linked to graphing or symbolic algebra tools. Other possible features of an electronic journal include searching and retrieving data on-the-fly, indexing, 'hyper glossaries', keyword searching and forward referencing. All but the latter are self evident — forward referencing is the concept that a document in the archive can be updated dynamically whenever a subsequent document refers back to it, so generating a hypertextual link 'forward in time'.

Netscape: The genesis of a new range of interlocked molecules

File Edit View Go Bookmarks Options Directory Window Help

Location: <http://www.elsevier.com/locate/chemcom/1996/1431-1430>

What's New What's Cool Destinations Net Search People Software

The Royal Society of Chemistry Chemical Communications, 1996, 1431-1430

The genesis of a new range of interlocked molecules

Peter T. Glink,^a Cesare Schiavo,^a J. Fraser Stoddart^{*a} and David J. Williams^b

^a School of Chemistry, The University of Birmingham, Edgbaston, Birmingham, UK B15 2TT

^b Department of Chemistry, Imperial College, South Kensington, London, UK SW7 2AT

Secondary alkylammonium ions self-assemble with suitably sized macrocyclic polyethers, such as the well known crown-24-crown-8 and 18-crown-6, into pseudocryptate superstructures in (a) solution, (b) the solid state and (c) the gas phase

Preamble

Self-assembly is a key concept in chemistry and physics, and is a key concept in the design and synthesis of large molecular assemblies and supramolecular arrays. The self-assembly paradigm relies upon the mutual recognition that exists between complementary molecules and ions through noncovalent bonding interactions. Adding to the repertoire of the known complementary molecules and ions are the 'supramolecular' chemist a greater choice in the number of different 'building blocks', or modules, with which to assemble increasingly more complex macroscopic materials. Thus, the identification of new systems which self-assemble spontaneously into supramolecular arrays of predetermined geometry is an important challenge in contemporary

Footnotes and References

Figure 2: Screen dump illustrating the feature article by Fraser Stoddart and co-workers. The article is displayed within the Netscape Navigator window. Underlined text is hypertextually linked to annotations which appear in the footnote frame. Molecular information in the article encoded in Protein DataBase (PDB) format appears as an animated in-line molecular model.

As an example of what can be achieved we have recently presented an enhanced version of a feature article taken from *Chem. Comm.* The paper by Fraser Stoddart and co-workers (Ref 15) describes recent advances in the area of the molecular design of interlocking molecules. In the electronic version figures in the paper version depicting molecular structures become animated 3-dimensional models using the Molecular Design Ltd *Chime*TM 'plug-in' for the Netscape browser (Ref 16). This product, based on the RasMol molecular viewer written by Roger Sayle (Ref 17), allows the reader to visualise complex structural information easily. Figure 2 depicts the article as it appears on the reader's screen. Note in particular the use of frames (a non-standard Netscape Navigator feature now included in the specification of HTML 3.2) to provide annotations. Other features in the demonstration article include hypertextual markup of spectral information and extensive cross-referencing. We have also used this technology to provide course notes, tutorials and workshops in a third year undergraduate module offered at Leeds on computer-aided drug design (Ref 18). The potential for the development of distributed courses from several participating universities and for open access to such materials should be obvious.

A relatively new development in browser technology are Java applets — small programs written in the Java programming language (Ref 19). In contrast to CGI scripts which run on the server, these programs are downloaded to the client and executed locally. Therefore they do not depend upon the speed of the connection to run. In some respects similar to the concept of plug-ins the advantage is that the user does not need to download and launch external applications in order to utilise Java, but unlike plug-ins a further advantage is that the Java code is a meta-language and hence platform independent. All that is required is a Java-compliant browser such as Netscape Navigator. Therefore the user needs no technical knowledge of the code and the publishers can incorporate any multimedia, animation or other such added value features which they desire, limited only by their coding ability and imagination. Java applets are relatively new technology: however, their immediate appeal has resulted in an explosion of interest and development, and already there are large numbers of excellent applets freely available on the Web. Our ongoing work with Java will be described at the meeting.

3. Conclusions

We have seen that the technology already exists, and is constantly developing, with which to create interactive documents which provide a multimedia experience for the user. There is no equivalent to this on the printed page, opening up many possibilities not only for the display and dissemination of research but also for teaching. The electronic medium will continue to have an increasing impact on the methods of, and attitudes towards, the publication of scholarly works. There is a certain amount of scepticism levelled at electronic publications because they blur the concept of what constitutes a 'publication'. In essence, however, a properly refereed electronic journal is no different from its paper equivalent. The risk is that until the scientific community at large agrees with this premise, to publish on the Web is to risk your work not being accepted as a *bona fide* publication. With all that this entails regarding research assessment exercises within academic institutions, etc., it is hardly surprising that some scientists may wish to publish on paper first. The level of mistrust can also be attributed in part to the proliferation of low quality information on the WWW: a criticism that can, however, also be levelled at the printed medium, for example some tabloid newspapers.

A properly refereed electronic journal, in particular one which has a respected printed version already in existence, has a good chance of becoming a respectable medium for publication. Whether electronic journals will continue to come out as purely parallel versions of their paper equivalents depends upon the attitudes of authors to this novel publishing medium. This in turn depends on the attitude of the scientific community in general. It is likely, however, that there will come a time when independent electronic journals are established and that the publications in such journals will be considered on a par with those of paper journals. After all, a journal's quality is determined by its refereeing process and content, not the medium by which it is presented.

From a publishers' perspective, the development of an SGML database has several obvious advantages, as outlined previously. The one major drawback is the question of maintaining an archive in the event of a particular publishing house ceasing to exist for whatever reason. There is no obligation on anyone to maintain the archive, and if there is no corresponding paper version, the information could be lost. This lends further weight to the feeling that paper journals could not be fully replaced by their electronic counterparts and that the two will exist in parallel. However it is probable that in the future libraries will redefine their function as we move into an electronic age and take on more of a role as primary archive repositories. At present, while publishers experiment with the Web, much information is provided free of charge. This situation cannot continue as more finance is ploughed into hardware, training of staff and so on and the journals go fully online. Publishers will have to develop a pricing structure. The problem is the level at which to charge. The electronic version of an established paper journal needs to be sufficiently different to allow the product to be sold at a price such that any extra production costs can be recovered. This makes added value crucial; the more extra features, the better chance you have of recovering your costs. In the case of a parallel electronic version of an established journal there is no extra cost for refereeing, but a purely electronic journal must address this issue in its pricing structure. This puts the independent electronic journal at an immediate disadvantage and tends to suggest that the <status quo> as regards the hierarchy of reputation of journals will be extended into the electronic form. This may cease to be the case, however, if the 'added value' of embedded data finds favour amongst the readership. In this case, and in our view this is a very real possibility, the future for purely electronic journals is bright.

The WWW opens up interesting possibilities for the collection of user feedback. A publisher can automatically

monitor, on a basic level, how many accesses individual articles have, but perhaps more interestingly how individual articles are read by the user. By experimenting with the format of a journal one could ascertain if users tend to look at images before the text, or vice versa, and in what order the article is read. Do readers access the abstract, conclusions and then the body of the article, or each section in order? This information is invaluable for publishers in order to make their product more attractive to the user, whether printed or electronic.

There are issues to be resolved regarding data misuse and the copyright of data. Models for dealing with copyright have not yet fully emerged and legislation to deal with this issue will probably be required. Data misuse is another issue, but abandoning the format in the face of what are in truth teething problems is not a sensible way to proceed. One can assume that, as with the technology, the law will advance, given time. There is little to be gained from waiting for it to catch up: the timescale of developments within the computing community is far shorter than that of the politicians and lawmakers, and the basic issues remain the same in spite of software developments.

Speed of access to the Internet is also a problem, as anyone who has tried to navigate the WWW following transatlantic links on a wet afternoon in the UK will readily testify. That such problems are encountered is evidence of the number of people using the system and it must be remembered how comparatively new this technology is. Precedent tells us that there is a reasonable expectation that such problems will be overcome in time. Apart from the obvious solution of building bigger and faster data highways, caching strategies and data compression techniques are already improving the performance of the WWW although these effects are often masked by the increase in usage. That said, the pace of development of the WWW is a testament to the ease with which the Internet allows information to be passed between individuals. A major criticism of the paper-based publication system is the timescale for work to appear: in some cases a year or more after the work was actually submitted for publication. The Internet, in particular in the form of discussion lists, allows scientists to exchange ideas freely and to keep abreast of rapid advancements, so avoiding the duplication of effort. There is a good chance that the use of the Internet will continue to spiral since there is a self-acceleratory aspect to it. The more people that use the Internet, the faster things develop, the more it has to offer, the more people it attracts and so on. From our experience the usage statistics on our Departmental server continue to show near exponential growth.

In summary, the impact of electronics on the publishing process will be enormous but constructive. The information at our disposal as scientists will be enriched by the advance of technology into this new medium for information dissemination. The fear that the printed journal will somehow be rendered obsolete is unfounded. In the near future all the evidence points to the electronic and paper publishing processes coexisting. In all likelihood they will follow parallel tracks, with the traditional printed form continuing to retain its place in the structure of information provision which exists currently. Finally, as the electronic format becomes more widely accepted as a respectable medium for publication, so publishers and authors will be able — and furthermore would be foolish not — to take full advantage of its unique features.

B.J. Whitaker
School of Chemistry
University of Leeds
Leeds
UK
E-mail: b.j.whitaker@chem.leeds.ac.uk

References

- [1] James, D., B.J. Whitaker, C. Hildyard, H.S. Rzepa, O. Casher, J.M. Goodman, D. Riddick and P. Murray-Rust (1995) *New Review of Information Networking*, **1**, 61.
- [2] *Report to the Joint Funding Councils' Libraries Review Group* (1993) Chairman: Prof. Sir Brian Follett, HEFCE, Bristol.
- [3] Knuth, D.E. (1970) *The TeX Book*, Addison-Wesley.
- [4] International Standards Organisation (1986) *ISO 8879: Information Processing — Text and Office Systems Standard Generalized Markup Language (SGML)*, Geneva.
- [5] Bryan, M. (1988) *SGML: An Author's guide*, Addison-Wesley.
- [6] J. English. CoST is available from <http://trystero.art.com/joe/cost/index.html>
- [7] J. Clark. sgmls is available from <ftp://ftp.jclark.com/pub/sgmls/>
- [8] A specification for HTML 2 can be found at <http://www.w3.org/pub/WWW/Markup/html-spec/html-pubtext.html>
- [9] A specification for HTML 3.2 can be found at <http://www.w3.org/pub/WWW/Markup/Wilbur/>
- [10] John W. Oster II (1993) Electronic Publishing Special Interest Group, ISO 12083.
- [11] Bachrach, S., P. Murray-Rust, H.S. Rzepa, and B.J. Whitaker (1996) *Network Science*, March. <http://www.awod.com/netsci/>
- [12] Rzepa, H.S., B.J. Whitaker and M.J. Winter (1994) *J. Chem. Soc. Chem. Comm.*, 1907.

- [13] Casher, O., G.K. Chandramohan, M.J. Hargreaves, C.J. Leach, P. Murray-Rust, H.S. Rzepa, R. Sayle and B.J. Whitaker (1995) *J. Chem. Soc. Perkin Trans. 2*, 7.
- [14] Pudiate, P.B. and B.J. Whitaker, work in progress. <http://cbl.leeds.ac.uk/~caybbp/>
- [15] Green, S. and B.J. Whitaker, work in progress. <http://chem.leeds.ac.uk/Project/Teaching/cadd.html>
- [16] *The Java Language: A White Paper* (1995) Sun Microsystems, Mountain View, CA.
- [17] Glink, P.T., C. Schiavo, J.F. Stoddart and D.J. Williams (1996) *J. Chem. Soc. Chem. Comm.*, 1483. http://www-clic.ch.cam.ac.uk/CLIC/html/c08349_1.htm
- [18] MDL Information Systems, Inc., San Leandro, CA, USA. See <http://www.mdli.com/mdlhome.html>
- [19] Sayle, R. (1994) *RasMol: A Molecular Visualization System*, Glaxo Group Research, Greenford, Middlesex, UK.



The Internet as an information source for environmental chemicals — first results of the evaluation of the meta-database of Internet resources

Kristina Voigt

GSF-Forschungszentrum für Umwelt und Gesundheit, Germany

Joachim Benz

Universität-Gesamthochschule Kassel, Germany

Rainer Brüggemann

Institut für Gewässerökologie und Binnenfischerei im Forschungsverbund Berlin, Germany

Abstract: An evaluation approach using the mathematical method of the Hasse diagram technique is applied on 20 environmental and chemical Internet resources. The data for this evaluation procedure are taken out of a metadatabase called DAIN (Metadatabase of Internet Resources for Environmental Chemicals) which is set up by the GSF Research Centre for Environment and Health and the University of Kassel. The URL for DAIN is <http://dino.wiz.uni-kassel.de/dain.html>. The following are chosen as evaluation criteria: search possibilities in Internet resources, quality of resources, number of chemicals in the resource, identification parameters for chemical substances, and information parameters for chemical substances. The 20 Internet resources are ranked with these five different evaluation criteria using a six-figure scoring system. A Hasse diagram is set up and discussed. A further analysis shows that the criterion 'information parameters for environmental chemicals' is the most important one in this ranking procedure.

Keywords: Internet, environmental databases, chemical databases, metadatabases, evaluation, Hasse diagram, ranking

1. Introduction

The Internet provides access to a vast variety of chemistry and environmental information. The quantity and the quality of these information resources continues to improve. However, there is an urgent need to help users of the Internet find the relevant information for a specific subject. A recently published article gives a brief guide to current online resources for environmental professionals (Ref 1). Chemistry resources on the Internet and chemical-specific requirements are given in several articles (Refs 2, 3, 4).

In chemical and environmental sciences several approaches exist to gather the relevant resources. These are lists of Web pages with or without short descriptions of the original site. In Table 1 a few internationally recognised lists of Web sites are given. This is only a small selection of what is available on the Internet. These listed resources are frequently used and recommended by many Webmasters. The table gives not only chemistry and environmental sites but also resources on toxicology and health, as well as sites treating the subject of material safety data sheets.

Table 1: Lists of environmental chemicals' resources on the Internet.

Name of the Resource	URL
Chemistry Information on the Internet	http://hackberry.chem.niu.edu/cheminf.html
Computer Chemistry Centre Erlangen	http://www.organik.uni-erlangen.de/
Directory of Environmental Resources on the Net	http://www.enviros.w.com/
Environmental Journalism	http://www.sej.org/
Environmental Sites on the Internet, Stockholm	http://www.lib.kth.se/lglg/eindex.htm
Fachgruppe Informatik im Umweltschutz	http://www.iai.fzk.de/Fachgruppe/GI/
FU CHEMnet WWW Entry Point	http://www.chemie.fu-berlin.de/index.html
Internet Chemistry Resources	http://www.rpi.edu/dept/chem/cheminfo/chemres.html
List of Environmental Toxicology Resources	http://pitt.edu/(martint/pages/envtox.htm
RSC Links to Other Chemistry Web Sites	http://www.worldserver.pipex.com/rsc/wwwsites.htm
Sheffield Chemdex	http://www.shelf.ac.uk/(chem/chemdex/
Where to Find MSDS on the Internet	http://www.chem.uky.edu/resources/msds.html
WWW Resources Chemistry	http://www.uky.edu/Subject/chemistry
WWW Resources Environment	http://www.uky.edu/Subject/environment
WWW Virtual Library, Chemistry	http://www.chem.uda.edu/chempointers.html
WWW Virtual Library, Environment	http://ecosys.drdr.virginia.edu/All.html#b

2. DAIN — Metadatabase of Internet Resources for Environmental Chemicals

An approach for a metadatabase called DAIN (Metadatabase of Internet Resources) was established in late 1995. DAIN can be found at the URL <http://dino.wiz.uni-kassel.de/dain.html>.

The set-up of the indexing-file system was presented at Online Information '95 (Ref 5). In this current paper the evaluation of the metadatabase which contains approximately 100 entries will be presented. During the evaluation approach those Internet resources where chemical substances are covered will be looked at. The data-field 'type of database' will be focused on. For the time being the following types of databases are covered: bibliographic, chemical catalogues, full text, inventories, metadatabases, numerical and structural databases. Numerical databases and metadatabases comprise the biggest proportion.

3. Relevant Internet resources for environmental chemicals and evaluation criteria

3.1. Selection of relevant resources

As the current number of resources is still quite easily comprehensible we selected 20 sites, mainly factual data sources. These are given in Table 6.

We searched all these resources again for the evaluation process and found out that only two sites had changed their location but clearly indicated their new address. This means that the chosen set of sites (also called objects in our evaluation approach) are well established data sources and will probably be available under their URLs for a longer period of time in the future.

3.2. Evaluation criteria for environmental chemicals Internet resources

Taking into account the different structure of Internet resources in comparison to online and CD-ROM databases (Refs 6,7) we developed two new evaluation criteria: 'search possibilities' and 'quality of Internet resources'.

3.2.1. Search possibilities

A variety of search possibilities exists for Internet resources. In this field of interest we came across the following: a simple listing of chemicals not even ordered alphabetically, an alphabetically ordered list of chemicals, and a search index for chemicals. Referring to the possibility of searching with the aid of a provided search form, some categories are encountered. The easiest form is one single search form, where the user types in for example the name of the chemical substance. More advanced systems provide several search forms with or without the possibilities of using Boolean operators. The most sophisticated search systems allow all three operators AND/OR/NOT to be used whereas most systems only allow AND/OR facilities.

According to our formerly introduced evaluation system for databases (Refs 8,9) we have worked out a six-figure scoring system where 0 is the worst score and 5 is the best one. Applying this scoring system to the current research activities concerning the evaluation of Internet resources, we set up Table 2.

Table 2: Scores for the criterion 'search possibilities in Internet resources' SE.

Criteria	Score
list	0
alphabetical list	1
index	2
search form, one field	3
search form, several fields	4
search form, several fields, Boolean operators	5

3.2.2. Quality of Internet resources

Data sources can be judged by many quality criteria. We have already published the relevant criteria in connection with the evaluation of online databases (Refs 7-9). However, in this paper we want to concentrate on quality criteria which we encountered while working with Internet resources. One important fact to know is whether the resource contains evaluated data. Expert knowledge is included in very few data sources. In contrast to other data sources, in the Internet world you find a number of awards which are given to some exceptionally good sites: for example the Magellan 3 and 4 StarSite, c/net Best of the Web and MetalWorker's Finalist TOP10 WebSite are mentioned here. The feature 'help' is not found very often within Internet resources. Therefore we also count this 'help function' as a quality criterion. Unfortunately not every site gives a short explanation on what the source is about, so the 'description of the resource' is another quality criterion.

It is quite obvious that more than one criterion may apply for one site. This is why we speak of a nominal criterion with respect to 'quality of the Internet resource'. Detailed explanations of nominal and ordinal criteria are given by Belke (Ref 10).

In order to come to a table of scores like that given in Table 2, four nominal criteria have to be interpreted in an ordinal manner which is now shown.

First the sequence of the criteria is defined:

evaluated data > award > help feature > description

Following the sequence the value 4 is given for evaluated data, 3 for award, 2 for help feature and 1 for description.

Then all possible combinations of these four criteria are calculated. The results are given in Table 3.

Table 3: Combinations for the criterion 'quality of Internet resource' QU.

Combinations	evaluated data	award	help feature	description	combination values
4 combination	x	x	x	x	10
3 combination	x	x	x		9
	x	x		x	8
	x		x	x	7
		x	x	x	6
2 combination	x	x			7
	x		x		6
	x			x	5
		x	x		5
		x		x	4
			x	x	3
1 combination	x				4
		x			3
			x		2
				x	1

Applying this combination method it so happens that for example the combination of three criteria (award, help feature, description) gives the same number as the combination of two criteria (evaluated data, help feature).

In order to come to the six figure method the combination values are divided by two. The scores are given in the Table 4. It has to be noted that this aggregation is not really necessary: it is performed in order to receive a homogenous treatment of criteria and to keep the list of criteria small.

Table 4: Scores for the criterion 'quality of Internet resource' QU.

Criteria	Score
none of the criteria	0
help function	1
description	1
help function, description	2
evaluated data	2
award	2
award, help function, description	3
evaluated data, help function, description	3
evaluated data, description	3
award, help function	3
evaluated data, help function, description	4
evaluated data, award, description	4
evaluated data, award	4
evaluated data, award, help function	5
evaluated data, award, help function, description	5

3.2.3. Other evaluation criteria

As we have been working in the field of evaluating online databases and CD-ROMs for some years, we have developed some criteria which are of great importance concerning the evaluation of environmental and chemical data sources. These are the 'number of chemicals' (NU), 'identification parameters for chemical substances' (ID)

and 'information parameters for environmental chemicals' (IP) which are also relevant for Internet resources.

The 'identification parameters' for environmental chemicals are structural formula, molecular formula, molecular weight, CAS number, other registry numbers, synonyms and so on. For the evaluation of environmental data sources, the criterion 'information parameters for environmental chemicals' is of great importance. It indicates what kind of parameters are covered in the resource in question. All these parameters have already been published (Refs 6,8,9). Whereas the already established scores for the criteria 'identification parameters' and 'information parameters for environmental chemicals' can be applied for Internet resources in the same way as for online databases and CD-ROMs, the 'number of chemicals' has to be adjusted to the topic of Internet resources. The number of chemicals in most of the Internet resources covered by DAIN is considerably lower than in online databases and CD-ROMs. Therefore we had to change the scores given for this criterion. The scores for 'number of chemicals in Internet resources' is given in Table 5.

Table 5: Scores for the criterion 'number of chemicals in Internet resources' NU.

Number of chemicals	Score
0-99	0
100-249	1
250-499	2
500-999	3
1000-9999	4
> 10,000	5

4. Evaluation approach using the Hasse diagram technique

Hasse diagrams have been described in several publications (Refs 11-13) and the repetition of the theoretical background and of other applications is beyond the scope of this article. However, for the sake of convenience to the reader some helpful remarks should be given: The basis of the Hasse diagram technique is the assumption that ranking can be performed while avoiding the use of an ordering index (Ref 13). Hasse diagrams visualise partially ordered sets. Partially ordered sets arise when an additional order relation is established between elements of a set. If for example Internet resources form a set then they are ordered if and only if all the properties of resource 'A' are equal or better than those of resource 'B'. In that case we write:

$$A \geq B.$$

The aforementioned order relation is an example among many possible realisations. By definition an order relation must be reflexive, asymmetric and transitive. An Internet resource may be compared with itself (reflexivity); if resource 'A' is 'better' than 'B' then the reverse is not true (asymmetry); and if resource 'A' is better than 'B' and 'B' is better than 'C' then it also follows that 'A' is better than 'C' (transitivity).

A graphical representation of such partially ordered sets draws a pair of ordered elements in the plane such that the better element is located above the worse one, and both are connected by a line. Lines which correspond to transitivity are omitted to keep a simple graphic.

Hasse diagrams are extremely useful if several criteria are given to decide which objects are priority objects. In this approach we look at 20 Internet resources (objects) which are evaluated with five different criteria using a six-figure scoring system (see Table 6).

Table 6: 20 selected Internet resources for evaluation, their criteria and scores.

Acc	Name of Internet Resource	SE	QU	NU	ID	IP
ARS	ARS PPD	1	3	2	3	3
ATS	ATSDR ToxFAQs	2	1	0	1	1
BIO	Biocatalysis/Biodegradation	3	4	0	4	2
CHE	Chemikalien Sicherheitsdaten	5	2	4	3	2
ECD	ECDIN	4	2	5	3	5
ENV	ENVIRO-NET	3	0	3	3	4
EPA	EPA CFS	1	0	2	2	4
EXT	EXTOXNET, PIPS	3	1	1	2	4
FIS	Fischer Scientific	4	2	4	3	2
HAC	Hazardous Chemical Db.	5	0	4	3	2
HAD	HazDat	3	3	1	2	1
ICS	ICSC	1	1	0	3	2
OPP	OPPT Chemical Fact Sheets	1	0	0	2	5
OXM	Oxford MSDS	2	0	4	2	4
PES	PESTIS	3	0	2	1	4
SID	SIDS	0	1	0	3	4
SIR	SIRI, MSDS Archive	3	1	4	2	4
STA	Stanford Uni Portfolio	5	0	4	2	1
UTM	Utah MSDS	2	0	4	2	4
WHO	WHO	0	1	0	1	2

4.1. Hasse diagram for 20 Internet resources

The Hasse diagram for this set of 20 objects is given in Figure 1.

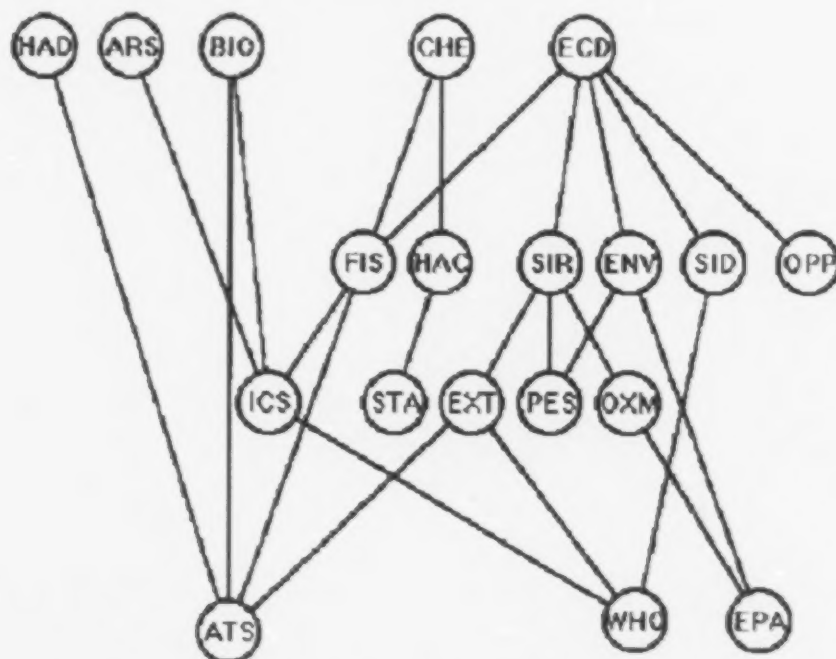


Figure 1: Hasse diagram for 20 Internet resources evaluated with five criteria.

The following conclusions can be drawn from the diagram.

The two resources OXM,UTM have the same scores (see also Table 6): they are members of a so-called non-trivial equivalence class. This means that these two Internet resources which treat the subject of material safety data sheets for chemical substances do not differ from one another if we only regard these five evaluation criteria explained in the above way. Five Internet resources are so-called maximal objects, which means that they are better regarding all five evaluation criteria than those objects which are situated in a lower level. The five objects are: ARS (ARS PPD), BIO (Biocatalysis/Biodegradation), CHE (Chemikalien Sicherheitsdaten), ECD (Environmental Chemicals Data and Information Network) and HAD (HazDat). These Internet resources are the best ones out of the chosen set of 20 sites. However, they are incomparable with each other. Taking, for example, a look at BIO (Biocatalysis/Biodegradation Database) and at ECD (Environmental Chemicals Data and Information Network) this can be demonstrated (see Table 6). BIO has a higher score in 'quality' identification parameters than ECD. The criteria 'search possibilities', 'number of chemicals' and 'information parameters' of BIO are less than those of ECD. Therefore no order relation exists, which means both Internet resources are not comparable. Consequently both resources are not connected by a line.

Minimal objects are those resources which have no successors. In Figure 1 the following six objects are minimal: ATS, EPA, OPP, PES, STA and WHO. These resources are the worst sites in this evaluation approach of a set of Internet resources for environmental chemicals. Taking a look at Table 6 it can be seen that these six minimal resources cannot be compared with each other (see explanation given for maximals).

Figure 1 shows four levels and six elements are found in the largest level.

Successors of a given resource, for example ARS, are those resources which can be reached beginning at ARS and following the lines downwards. This means the successors of ARS are ICS and WHO. With respect to all the given evaluation criteria in Table 6, these sites are worse than ARS.

Taking a look at the successors of the maximal object ECD, it can be seen that ECD has 13 resources which are situated lower in the Hasse diagram. This means that ECD has a considerably higher number of objects which have lower scores than for example CHE, BIO, ARS and HAD. For example HAD is only comparable with ATS.

4.2. Sensitivity analysis of ranking with respect to its evaluation criteria

The ranking of a set of objects does not only depend on the numerical values (scores) but even more on the choice of criteria. In other words the ranking of objects is sensitive to the set of criteria. A matrix called the W-matrix which has been discussed by Brüggemann (Ref 14) quantifies the influence of the evaluation criteria on the ranking. In case 0 all evaluation criteria — that is to say in this example all five evaluation criteria SE, QU, AN, ID and IP — are looked upon. If one criterion out of five criteria is omitted then $n+1$ cases are possible. For this evaluation six cases result. In case 0 all criteria are evaluated (see Figure 1). In the following five cases one criterion is omitted at a time according to Table 7.

Table 7: Cases for five evaluation criteria.

Case /Criteria	SE	QU	AN	ID	IP
Case 0	x	x	x	x	x
Case 1	-	x	x	x	x
Case 2	x	-	x	x	x
Case 3	x	x	-	x	x
Case 4	x	x	x	-	x
Case 5	x	x	x	x	-

The Hasse software which was developed by the third author (Ref 14) is able to calculate the W-matrix for all objects. This calculation for the five criteria gives the following results:

- Case 1: (SE omitted) 18 changes in the Hasse diagram
- Case 2: (QU omitted) 29 changes in the Hasse diagram
- Case 3: (AN omitted) 9 changes in the Hasse diagram
- Case 4: (ID omitted) 6 changes in the Hasse diagram
- Case 5: (IP omitted) 46 changes in the Hasse diagram

The more changes that are induced by omission of a criterion, the more important the criterion in question is.

This means that in this evaluation approach 'information parameters for environmental chemicals' is the most important criterion followed by 'quality of Internet resources' and 'search possibilities in Internet resources'. The two criteria which are set-up especially for Internet resources have a big influence on the ranking of the 20 chosen Internet sites.

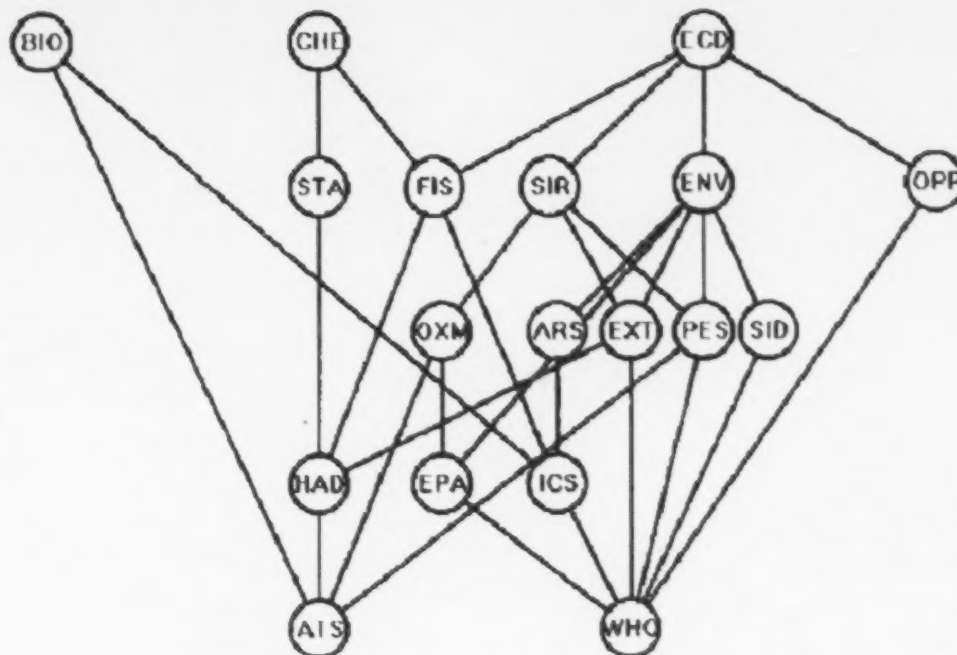


Figure 2: Hasse diagram for Case 2 (Omission of QU): 20 Internet resources.

Comparing this diagram with only four evaluation criteria with the diagram in Figure 1, a few changes can easily be seen. It is obvious that only three instead of five maximal objects are existing now. HAD and ARS are respectively in the fourth and third levels of the diagram. This Case 2 diagram has five levels as opposed to four in the Case 0 diagram. In this diagram only two resources, ATS and WHO, are minimals. CHE/HAC and OXM/UTM are members of non-trivial equivalence classes.

5. Future research

Apart from the content of the Internet resource the update plays an important if not vital role. In this respect two aspects of the update of a resource have to be examined. First it is important to check if the resource can still be found under the same URL and second one has to figure out if and how the resource has been updated. Automated systems to fulfil this urgent need to inspect the update of an Internet resource are under development.

Additionally a great effort will be made to extend the DAIN Metadatabase of Internet Resources for Environmental Chemicals continuously. The authors would be very grateful for proposals and hints in this regard.

Furthermore the Hasse diagram technique will be extended by additional features coming from:

- graph theory (for example identification of articulation points as points of interest for further data analysis);
- multivariate statistical methods, for example integration of cluster analysis;
- establishing more instruments to define the similarities of Hasse diagrams.

Kristina Voigt
 GSF-Forschungszentrum für Umwelt und Gesundheit
 Projektgruppe Umweltgefährdungspotentiale von Chemikalien
 Ingolstaedter Landstrasse 1
 85758 Oberschleissheim
 Germany
 Tel: +49 89/3187 4029
 Fax: +49 89/3187 3369
 E-mail: voigt@wiz.uni-kassel.de

References

- [1] Reichardt, T. (1996) Environment On-Line — a Guide to Internet Resources, *Environ. Sci. Technol.*, **30**(2), 76–81.
- [2] Ihlenfeldt, W.-D. and J. Gasteiger (1995) Chemie auf dem Internet, Neue Wege der Vermittlung chemischer Information, *Chemie in unserer Zeit*, **29**(5), 249–259.
- [3] Kazic, T. (1995) Chemical information: how do we get it and what do we do with it? in H. Collier (Ed.), *Proceedings of the 1995 International Chemical Information Conference*, Nimes, France, 23–25 October 1995, Infonortics Ltd, Calne, pp. 48–61.
- [4] Rzepa, H.S., B.J. Whitaker and M.J. Winter (1995) Chemistry and the WWW, *Chemistry in Britain*, 685.
- [5] Benz J. and K. Voigt (1995) Indexing file system for the set-up of metadatabanks in environmental sciences on the Internet, in D. Raitt and B. Jeapes (Eds.) *Proceedings of the 19th International Online Information Meeting*, 5–7 December 1995, London, Learned Information Ltd., Oxford, pp. 455–466.
- [6] Voigt, K. and R. Brüggemann (1994) Survey of environmental databases, comparison of search results, in D. Raitt and B. Jeapes (Eds.) *Proceedings 18th International Online Information Meeting*, 6–8 December 1994, London, Learned Information Ltd, Oxford, pp. 159–172.
- [7] Voigt, K. and R. Brüggemann (1994), Evaluierung von Umwelt-relevanten CD-ROMs, in L.M. Hilty, A. Jaeschke, B. Page and A. Schwabl (Eds.) *Informatik für den Umweltschutz*, 8, Symposium Hamburg 1994, Metropolis-Verlag, Marburg, pp. 187–194.
- [8] Voigt, K. and R. Brüggemann (1994), Ansatz eines Scoring-Systems für chemische/umweltrelevante Online Datenbanken, in C. Jochum *Software-Entwicklung in der Chemie 8*, GDCh-Fachgruppe Chemie-Information-ComputerSoftware-Entwicklung in der Chemie 8, *Proceedings of the 8th Workshop 'Computer in Chemistry'*, Seeheim-Jugenheim/Darmstadt, 17–19 November 1993, Gesellschaft Deutscher Chemiker (GDCh), Frankfurt/Main, pp. 1–15.
- [9] Voigt, K. and R. Brüggemann (1994) Scoring approach for environmental databases, in *Proceedings of the National Online Meeting*, New York, 10–12 May 1994, Learned Information Inc., Medford, NJ, pp. 427–439.
- [10] Belke, W., D. Graichen and M. Starruss (1979) *Nichtmetrische Klassifizierung von Informationen, Theorie und Anwendung*, Akademie-Verlag, Berlin.
- [11] Brüggemann, R., K. Voigt and E. Halfon (1996) Partially ordered sets — a computerized tool to compare environmental databases, in R. Denzer, G. Schimak and D. Russell (Eds.) *Environmental Software Systems, Proceedings of the International Symposium on Environmental Software Systems 1995*, Chapman & Hall, London, pp. 185–195.
- [12] Davey, B.A. and H.A. Priestley (1990) *Introduction to Lattice Theory and Order*, Cambridge University Press, Cambridge.
- [13] Halfon, E. and M.G. Reggiani (1986) On ranking chemicals for environmental hazard, *Environ.ScienceTech.*, **23**, 1173–11679.
- [14] Brüggemann, R. and E. Halfon (1995) Theoretical base of the program 'Hasse', GSF-Bericht 20/95, GSF-Forschungszentrum für Umwelt und Gesundheit, Neuherberg.



Ian Ritchie

British Computer Society, UK

1. Overview of the UK media industry

The media industry is the fastest growing business in the world today; additional leisure time, coupled with increasingly global distribution, has created large international markets for information and entertainment. The UK is relatively strong in the three main areas concerned with new media publishing: IT, telecommunications and publishing. The revenue of the telecommunications and IT industries in the UK is over £30 billion, similar in scale to that of healthcare. The UK broadcasting and publishing industries each stand at approximately £5 billion, representing around 10% of the world market.

Technology deployment in the UK is strong with a competitive telecommunications industry and widespread deployment of digital and optical fibre technology. The extent to which the UK consumer is willing to adopt new technology is also strong, indicated by the growth in satellite TV (the number of BSkyB subscribers having grown from 243,000 to 3,478,000 between 1989 and end-1994). One hundred and twenty-five franchises for cable television have been awarded and by mid-1995 4.7 million homes were capable of connection, of which approximately 900,000 are active customers. The current franchises awarded, when completed, have the potential to reach approximately 70% of all homes.

All of these factors would seem to indicate that the UK will continue to be an excellent breeding ground for the development of new media industries. The UK remains particularly strong in areas such as news gathering and distribution, television production and publishing, music production and publishing, and interactive games development.

2. New media publishing areas

New media technologies are being developed in a number of complementary and competing areas. It is interesting to compare and contrast such developments in the context of the knowledge that digital techniques for production, recording, archiving, distribution and delivery are fast becoming the standard.

This section reports on the state of publishing in several fields: corporate publishing; news and periodical publishing; scientific, technical and medical publishing; teletext and videotext; CD-ROM; and platform games. Although these fields seem at first sight to be quite distinct from each other it is likely that the distinctions between them will begin to blur during the coming decade.

2.1. Corporate publishing

In a typical large corporation, technical and procedural documentation line the walls of offices and maintenance workshops. Much of this material will almost certainly be outdated and most of it will never be read. While the need for information is rapidly expanding, printed documentation is proving ineffective when it comes to delivering large amounts of information at the point and time of need. In addition, printed matter can be extremely expensive to produce and distribute.

An average company or department of 500 workers generates over 1000 pages of new or updated policy, procedural and operational materials each year. An author needs on average six hours to compose each page (valued at, say, £200 per page), which will be passed along for review and returned for revision. Once written, each page costs the organisation more than £40 to print and distribute and another £10 each time it needs to be updated. Between publishing and revision, there are the added clerical costs of filing, storing and maintaining paper documents. Each year, the typical organisation with 500 employees can spend well over half a million pounds on paper-based information flow.

Unlike a printed document where structural elements such as chapter headings are merely signposts, electronic hypermedia documents can allow readers to use these elements to zoom in or out of a chapter to greater or lesser levels of detail. For example, a reader can target a chapter heading to open that chapter and scan its list of subheads, then focus in on one of the subheads to expand and display its contents. A reader can also search through an electronic document or publication by following links created by the author to represent relationships between topics. Rather than being structural or hierarchical, these relational links connect topics that you might see cross-referenced in a book's index.

Increasingly such sets of electronic documents are being stored on a file server, accessible and downloadable over a computer network or recorded on a low-cost CD-ROM. Readers can find what they need without having physically to leave their tasks. If they have laptop computers, readers can access information whether they're at

their desks, on the shop floor or at a customer site.

The creation of interactive electronic publishing has greatly increased the effectiveness of many organisations. As in many other fields of technology, corporate publishing has demonstrated what can be achieved, and the experience in this field is proving very helpful as these techniques become applied to other areas. These needs, along with other more general communication requirements have led to recent growth in 'intranet' or private closed services using Internet protocols but not connected outwith the organisation.

2.2. News and periodical publishing,

The newspaper and magazine market is an extraordinarily diverse one, where some mass-market daily newspapers (such as the *Sun*) sell 4 million copies every day and other expensively produced, well circulated and well regarded titles (such as the *Financial Times*) sell only a fraction of that. Each product, whether a newspaper or magazine, exists on a combination of revenue from subscription and advertising and the mix of these is crucial to the success of individual publications. All newspapers and magazines are subject to enormous competitive pressure and, in many cases, are experiencing long-term market decline. Goldman Sachs estimates that the total market size for newspapers worldwide was \$137bn in 1993 and they believe that there has been a general 12% decline in market size over the preceding four years.

Mass market newspapers are relatively dependent on their cover price. Due to the fact that they have a very wide circulation most of their advertising is display advertisements for mass-market products and national retail chains. Other newspapers such as the national broadsheets, regional titles and specialist magazines can offer a more targeted audience and, as a result, command a higher relative value of advertising income, particularly classified advertising such as job recruitment and real estate.

The decline of the newspaper has been long forecast but the product has been very robust. New titles, such as *The Independent* in the UK and *USA Today* in the US, have shown that it has still been possible to launch new products into the market. One area of decline however has been in that of regional titles, particularly evening newspapers where most have lost circulation, year on year, for the last 20 years.

Specialist magazines have enjoyed a new lease of life following the development of desktop publishing (DTP) in the late 1980s. DTP has slashed the production costs of publications and has allowed many more titles to be developed economically than ever before. The sale of such titles shows a desire by consumers to obtain information specifically focused on their interests — a desire which is appropriate for emerging Internet publishing systems where specific interest groups can be easily served.

New production techniques have allowed products to be distributed more effectively. Even in a small country such as the UK all national newspapers are edited and typeset in one location and then printed in several satellite printing plants. Titles such as the *Financial Times* and the *Wall Street Journal* are printed simultaneously in many print plants all over the world, as are topical magazines such as *Newsweek*, *Economist*, *New Scientist* and *Time*.

The arrival of digital publishing over the Internet, particularly using commercial online networks and the World Wide Web, provides such publishers with potentially vast new outlets for their editorial content. A study conducted by American Opinion Research found that 19% of newspaper editors and publishers believe online services will become their biggest competitors within five years. Almost half were planning some online service version of their publication.

Dozens of newspapers are arriving on the Internet every month, from major international titles such as the *Financial Times*, the *New York Times*, the *Daily Telegraph* and the *Wall Street Journal* through leading regional titles such as the *San Jose Mercury*, through to small local newspapers. Most of these, however, are best described as 'work in progress', where the final mix of subscription and advertising has yet to be established.

The *Daily Telegraph*, which launched an edition on the Internet in November 1994, gets 10,000 accesses and 1000 e-mails a day. By asking users to register before they can view the newspages it can gather statistics about its readers and this evidence can be used to justify advertising rates. The Electric Telegraph is typical of many Internet newspapers projects in that:

- a small staff have been set up to repurpose existing editorial material from the newspaper to online availability;
- no fee is being charged for access;
- advertising is currently attracted but is much more limited than the print edition;
- the product is not editorially distinctive from its parent publication.

A few newspapers, such as the *San Jose Mercury* and *USA Today*, have tried to set up subscription charging but it seems unlikely at the moment that significant proportions of the target audience will pay extra for general interest information, particularly since much of their competition remains available for free. After originally attempting to charge a subscription of \$12.95 per month, *USA Today* relented and have abandoned their subscription charge for the present. The *San Jose Mercury* is charging a subscription to access a very localised area of its site and this effort seems to be largely aimed at cross-marketing with their regular subscribers.

An example of the difficulty of introducing subscription is shown by the *Colorado Springs Gazette-Telegraph* which tried to sell access for \$9 per month, or \$5 per month if you had an Internet account with their local access provider. Visitors dropped from 300-500 a day when it was free to practically nil when they began charging for access.

The Electric Telegraph is currently available free but is considering whether to introduce a subscription charge, as are other UK newspapers such as the Glasgow-based *Herald* and *Times* — a mixture of material

derived from the morning *Herald* and the *Evening Times*.

The most ambitious development in the UK however has been Times Newspapers, which has put the daily *Times* and the *Sunday Times* on the World Wide Web for free access. There is no sign that they plan to charge a subscription but this, like their recent price war in their printed titles, may primarily be a tactic to ensure that their competitors do not gain unfettered access to a new source of revenue.

Another approach, used by the *Economist* in its d.Conn online publication and the Cnet technology cable channel, is to create wholly new titles which are directed at a specific target audience and are only available via the Internet. By being more specialist, there is perhaps a better prospect of charging a subscription fee in the future, or advertising targeted at their specialist audience.

After years of experimenting with a limited product on AOL the *New York Times* launched a full version of the newspaper on the World Wide Web in January 1996. Their pricing is unusual in that it is free in the USA but \$35 per month elsewhere. Since, on the Internet 'nobody knows you are a dog' it is difficult to see how the *New York Times* knows whether or not you are an American!



Figure 1: The New York Times 'electronic front page'.

The *Wall Street Journal*, after experimenting for nearly a year with a limited product, is now publishing a full online product on the World Wide Web. They have set an annual subscription at a highly attractive price of under \$50. Both the *New York Times* and the *Wall Street Journal* are clearly attracting a steady stream of advertising and it seems likely that these publications are already contributing significant revenue to their owners.

One key question for interactive publishing projects is this — when users are able to select freely from any publication around the world, what will make them choose yours? The generation of some form of loyal user base, or 'community', will be vital.

We will probably also see the splitting of editorial and classified advertising, creating difficulties for many existing publishing markets. Currently the business model of newspapers such as the *Daily Telegraph*, *Guardian* etc. is heavily dependent on delivering classified advertisements effectively to a key audience. The newspaper is a complete package which for its purchase price delivers a combination of editorial and specialist advertising.

It is highly likely, however, that a batch of free World Wide Web 'publications' will begin to appear funded entirely on classified advertisements, each directed to a specific audience. Online advertisements are capable of being much more focused on the needs of the user and can cheaply include colour photographs, audio files etc. In this context it may be a significant fact that in the UK local newspaper market there are more 'freesheets' than local paid-for newspapers.

If classified advertising starts to move to online delivery it will have a devastating effect on many existing publishing franchises, such as broadsheet newspapers and the technical and trade press.

2.3. Scientific, technical and medical publishing,

To a large extent the scientific, technical and medical (STM) publishing market is, or has the potential to be, ahead of more general markets in adopting electronic publishing since the academic and research community are almost entirely 'wired' up to the Internet already. For a number of reasons, mostly due to commercial inertia and the inherent conservatism of a field where the medium of publication (ideally a highly regarded peer reviewed journal) carries almost as much value as the content of the material itself, this has not been the case.

Traditional STM publishing is the means by which scientific endeavour becomes promulgated, challenged and further developed. Researchers submit papers describing the results of an area of research to a suitably respected journal. The editors of the journal submit it to a number of referees, experts in the field, and after satisfying the referees that the material is sound it is printed in the journal. The readers of a journal are normally the same community as the authors.

STM publishing companies such as Oxford University Press, Elsevier and Springer-Verlag provide added value to this process by ensuring the quality of the editorial, production and marketing of the journals. The subscribers for such journals are often the technical library of a university or the research department of a company. As the customer is not an individual it has become possible to charge relatively high prices for the journals, well in excess of their production costs, much of which is required to recoup the marketing expenses. The effect of this is that major established journals with relatively large circulations have become highly profitable.

This system is now beginning to break down. To save money, libraries are challenging their users to cancel all but the most necessary journals, and online researching systems such as Blackwell's Uncover allow working researchers to identify specific published papers and have them retrieved, on a per copy basis, from central copyright libraries, eliminating the need to hold the copies locally.

Academics are also challenging the need for publication on paper, given the costs both in time and money. By using the Internet to publish their results, researchers can eliminate most of the expensive parts (typesetting, printing, warehousing, distribution and shipping) of the publication process. The other essential components, editorial quality and marketing, may not necessarily need the services of a traditional STM publishing company. Editorial quality has always crucially depended on the academic community in any case. As regards marketing, most technical fields have established extensive channels of communication via professional and learned societies, newsgroups, and listservs. News about a new form of publication can be made available to such a group very effectively.

An early example of this new form of by-passing STM publishing is HEP, a 'pre-print' archive of papers in high energy physics run by Paul Ginsparg at Los Alamos laboratory. This project began in August 1991 and has since gathered almost all the papers published in the field of high energy physics worldwide, submitted by the authors in 'pre-press' form but in most cases revised when the final papers are published in traditional journals. (HEP can be reached via the World Wide Web at <http://xxx.lanl.gov/> and is receiving accesses from around the world at the rate of 35,000 a day). Currently this system co-exists with the printed journals but increasingly the HEP archive will become the natural resource for access to research information in its field, calling into question even the need for the paper versions to be produced.

This kind of development has major implications for the STM publishers and most are beginning to experiment with electronic publications, such as Springer-Verlag's *Journal of Computer Science* which is proposed as a simultaneous electronic/paper product. Due to the lack of current charging systems capable of collecting small amounts of money these systems are implemented by subscription systems which are accessed by password control.

2.4. Teletext and videotext

There is a tendency to assume that online services are new but in the early 1980s several online services were developed both in Europe and the USA, based on a connection between a terminal connected to a telephone line and a television or computer as a display. Several large publishing, telecommunications and media companies launched services, most of which lost a great deal of money and were closed. In the USA these included Gateway, from Times Mirror, which was killed in 1986 after an estimated \$20 million investment and a system from Time Inc., launched in 1982 in San Diego and Orlando, which was shut down after one year and an investment of more than \$25 million.

In the UK the Post Office Telecommunications (now BT) marketed Prestel in the late 1970s which, apart from being widely adopted in the UK travel services industry, has never achieved its potential and has been scaled back.

A variant of videotext, called teletext, uses spare capacity in the broadcast television signal to send data along with the broadcasts. This is now built in to most televisions sold in Europe and is supported by services of news, information and listings, often supported by advertising and commercial data. In the UK this service is called Ceefax (BBC) and Oracle (ITV). It is broadcast-only and users must wait for the desired 'pages' to cycle round before loading into their television.

A new teletext-like initiative called InterCast was announced in October 1995 by an alliance of 13 companies including Intel, NBC, CNN, Viacom, WGBH, QVC, Comcast, America Online, Asymetrix, En Technology and Netscape. It allows HTML encoded documents to be broadcast embedded in the broadcast signal. Suitably equipped personal computers will be able to receive such broadcasts and assemble electronic documents, which might be of general interest, or might be used to add value to a specific television broadcast (<http://www.intercast.org/>). A similar document broadcasting system is currently being planned for Europe by a

Online Information 96 Proceedings

partnership between BT and BSkyB.

The undoubted success story in the world of videotext is Minitel in France, where the terminals were originally given to telephone subscribers in lieu of a telephone directory.

Minitel uses a small, easy-to-use terminal with a keyboard and a screen using a character-based display with simple graphics. The service operates at 1200 baud and gives access to more than 25,000 online business, leisure, financial and information services. A total of 6.5 million Minitel units have been installed, two-thirds of which are used by households, and in the last year over 1 billion Minitel calls were made.

France Télécom collects transaction fees on behalf of information providers and bills the customer on their telephone bill: last year \$605 million was paid in this way to information suppliers. France Télécom continues to develop the Minitel service: it has raised the speed of the service to 9600 baud for the 'Teletel Vitesse Rapide' variant and ISDN at 64 kb/s is also being provided in the form of 'Teletel Acces Numeris'.

2.5. CD-ROM

From a slow start (players for PCs have been available for over 10 years) CD-ROMs are now becoming a normal part of the configuration of most PCs sold. During 1994 54 million CD-ROMs were sold (according to Dataquest) and the installation of CD-ROM players is one of the fastest growing areas in the personal computer market. There are two distinct types of CD-ROM publication, corporate and consumer.

2.5.1. Corporate and database CD-ROMs

The CD-ROM, which can hold approximately 650 Mb of storage for a per-copy cost of well under £1, has been seen to be ideal for publishing large archives of technical, financial and reference materials. Companies such as Dun & Bradstreet and KR OnDisc publish such products but the largest player in this field is SilverPlatter, which has over 200 current titles on subjects such as Agriculture Research, Cumulative Book Index, Health and Drug Information Library and Petersen's College Database. Most of the SilverPlatter titles are co-published with an existing specialist publisher: in most cases the content has already been assembled and the conversion costs of moving the information to CD-ROM, including the addition of searching software and the user interface, is relatively low. They can also usually be marketed through the same channels and in cooperation with the co-publisher who will already have access to special interest groups.

2.5.2. Consumer CD-ROMs

As a contrast to corporate and database publications, very few consumer CD-ROM publishing companies are currently making much money. The problem lies with the economics of multimedia development and with the difficulties of reaching the customers through the distribution channels. It costs typically £150,000 or more to develop a reasonably professional CD-ROM title, and even more if third party rights have to be negotiated. A CD which sells for £30 at retail will return £12 or so in revenue to the publisher (after the distributor and dealer have received their margin) and, based on a typical royalty rate of 10% of net income, the developer can expect to get approximately £1.20 for each copy sold. It is therefore necessary to sell well over 100,000 copies in order to recoup the development costs alone. Book publishers don't have this kind of problem — a book is normally written by a single individual, often using discretionary time, and is edited part-time by another person. The cost of developing a title is therefore only a few thousand pounds. If a book turns out to be a success everybody starts to make money very quickly. When a CD-ROM is a success — and 250,000 copies is currently the kind of sales figures which would be regarded as a considerable success — the profit margins are still not huge. Many of the CD-ROM publishing companies which expanded in the boom times of 1994 have now been forced to cut back. For example, one of the earliest and well-regarded publishers in this field, Compton's New Media, has recently reduced its product range from over 300 to 60 and has laid off 30% of its workforce.

2.6. Platform games

A very large publishing market which has emerged in recent years is that of platform games, dominated by the manufacturers Nintendo and Sega and recently joined by Sony. These manufacturers sell special-purpose graphics computers without a keyboard which use a television set as a display device and have a cartridge interface which can be used to install individual programmes. These cartridges, which typically cost £50 or more at retail, will contain one game stored on computer read-only memory. Most of these games are highly visual and interactive and have high quality animation and sound effects.

Platform games are a 'hit parade' business with the top 10 games making a huge profit. Last year's top selling game, 'Donkey Kong Country' developed by Rare Ltd in Warwickshire, sold over 7.5 million copies in the six months following its launch in November 1994. This single game therefore represents a market that is comparable in size to the entire global wordprocessing market for personal computers (or about half the size of the UK commercial television industry). Rare recently announced that it had sold a 25% equity stake to Nintendo and that the proceeds would be used to increase its staff from 84 to 250 in the next two years.

One characteristic of platform games is that the manufacturers, Nintendo and Sega, retain a great deal of control over the software and charge a royalty on all games sold. Another characteristic is the furious pace of development: Nintendo and Sega have regularly upgraded their products capabilities and have been challenged by CD-ROM based platforms such as the Sony Playstation, Philips' CD-i and 3DO which has been brought to market by Panasonic. They are retaliating by launching upgraded platforms themselves: Sega with the new CD-

ROM based Saturn platform and Nintendo with its forthcoming Ultra64.

One recent development in the world of platform games is an added Internet capability. Both CD-i and Sega Saturn have announced add-on kits which will turn their platforms into Internet terminals.

The Nintendo Ultra64, due for release in the autumn of 1996, is a remarkable feat of practical mass market technology. It will contain much of the computer graphics display power of a top of the range Silicon Graphics workstation (currently priced at over \$100,000). The Nintendo Ultra64 on the other hand will sell for \$250 or so when it reaches the shops.

Development of titles for platform games has up to now been a fairly specialised task. But increasingly development tools are being made available which allow developers to create a product for a number of different formats, including CD-ROM for personal computers.

3. Convergence

The next decade will see remarkable development and convergence in all of these markets. These developments will be based around three distinct areas where emerging technologies will offer a dramatic improvement over current systems:

- high bandwidth, low-cost communications channels;
- low-cost, high-performance terminals (so called Network Computers);
- easy methods of collecting payments, including small quantities of payments.

Telecommunications operators, cable television companies and satellite broadcasters are aiming to provide high bandwidth connections by deploying new technologies such as ADSL, ATM and localised or satellite broadcast high-frequency radio. It is quite likely that such delivery channels will enable a range of attractive consumer products to be developed and sold, such as video on demand, and that on the back of such channels it will be possible to deploy highly sophisticated multimedia publishing products.

The arrival of the Network Computer will allow a low cost, easy-to-use device to be marketed for domestic use which it is hoped will be less technologically intimidating than the modern multimedia personal computer. Several companies such as Oracle and Acorn have already announced product and many of the platform games manufacturers will undoubtedly see this as their natural market.

The growth of the Internet and the potential development of Network Computers could create an environment which will encourage a great deal of convergence. Existing publishers, from television and radio broadcasters, newspaper and magazine publishers through to games developers could all gain a common new mass-market distribution method and potentially access to a fast-growing market. Organisations as diverse as BT, the BBC and Virgin have recently announced online entertainment and information products which will be launched soon.

One of the main reasons that Internet-based new media publishing has not taken off commercially is that it is currently very difficult to collect money for your product. Several banks and financial services companies have been developing electronic payment systems and it is now very likely that low-cost payment systems will soon arrive which will allow anyone to freely buy and sell material over the Internet.

We are poised at the beginning of the development of an exciting new industry, which will transform the way in which we will communicate with each other during the 21st century.



Creating new pricing models for electronic publishing

David B. Boelio and Nancy H. Knight

CD-MAX Inc., USA

Abstract: *The pricing of electronic information is a key element to its market success. Since the introduction of electronic information various models have been used to price information services. This paper provides a historical context for the discussion of pricing issues, describes the special pricing challenges confronting providers of electronic information, presents alternatives to the fixed pricing models most commonly used today and reviews several examples of new approaches to pricing made possible by technological advances.*

Keywords: unit pricing, fixed pricing, usage-based pricing, pricing models, cd-rom, internet, online

1. Introduction

Establishing pricing policies that are fair and flexible is of vital importance to the information industry. For information providers, pricing is a key element of competitive positioning, market penetration and profitability; for information users, pricing represents a principal determinant of access to information. Commercial enterprises necessarily charge 'what the market will bear,' a catch phrase that implies both efficiency and sensitivity to customer needs in the formulation of prices.

The pricing of most information available electronically is far less efficient and market-sensitive than it could be, however. This paper will provide a historical context for the discussion of pricing issues, describe the special pricing challenges that confront providers of electronic information, present alternatives to the fixed pricing models most commonly used today and review several examples of new approaches to pricing made possible by technological advances.

2. History

From the advent of electronic information, information providers have been challenged with pricing issues. Pricing metrics have been dictated by a combination of what is familiar and what is technologically possible. Before electronic information was available, information providers (referred to as publishers in those days) sold printed information by both a unit price metric, for example books (outright purchase), and journal subscription (annual fee per year). Purchasers 'owned' the printed materials they bought: if they discontinued a subscription, they kept what they subscribed to at that point.

These pricing metrics are predictable and consequently budgetable. The downside of printed information is that for some fields it is quickly out of date and requires large amounts of shelf storage space.

With the advent of online and the capacity to store vast amounts of information on mainframe computers, currency and storage issues were solved. However, the unit pricing metrics no longer fit. Publishers of electronic information were challenged with the tremendous costs of creating and maintaining electronic files, and vendors were challenged with the costs of developing search systems, loading and warehousing large files. In addition, there were telephone line charges to connect users with the mainframe computers. As a result, a usage-based pricing metric based on connect time was born. Users paid for the amount of time they were connected to the mainframe, plus additional charges for printing out specific information.

Usage pricing based on connect time solved some pricing issues: it provided a means to compensate information providers and vendors for using current information. It also created a new market for online training. Since pricing was based on connect time, it behoved the searcher to be an 'expert,' knowing all the right commands. Being an accurate and fast typist was also key. What the connect time metric did not solve were the issues of predictability and budgetability. It also necessitated an 'intermediary' who knew all 'the tricks' of the online system to keep costs to a minimum. Frequently this resulted in the actual end-user being a step removed from the information.

When standalone CD-ROMs entered the scene with their fixed (subscription) price metric, it solved some of the disadvantages of the online connect hour price metric. It provided unlimited searching for a fixed fee which encouraged high use. Search systems were designed for the end-user, removing the need for an intermediary to

do the search, and the length of time to perform a comprehensive search was no longer an issue (unless, of course, there was a queue of users waiting to use the same CD-ROM). The downside was establishing the 'right' subscription fee — high enough to compensate information providers and vendors and low enough to attract subscribers.

Enter the networked CD-ROM. The pricing metric was again re-visited by many information providers and, in general, variations of the fixed pricing method held. (See discussion below in Section 5, 'Discounts for high volume users'.)

With a strong tradition of fixed pricing resulting from CD-ROM, it is not surprising that many of the first Internet information offerings are based on fixed pricing, providing unlimited searching over a period of time for a fixed amount of money.

3. Why new pricing models are needed

The same technological advances that create new ways to package, distribute, use and sell information also create pricing dilemmas for providers. Powerful indexing and search tools allow information users to take advantage of the granularity of data in digital form by retrieving from large databases only those items that meet their search criteria. Records, once retrieved, can be tailored for use at will by means of copy and cut-and-paste commands, assuming no controls are in place. Customers desire and use information in discrete amounts. Yet when providers offer it in bulk, as is usually the case with CD-ROM databases, they risk creating unfulfilled demand and losing sales.

Pricing issues also have a security and copyright dimension. Providers are concerned about losing control over high-value electronic information that can be replicated an infinite number of times, accessed by and sent to thousands of potential users, and stored by any or all of them at a negligible cost. Inflexible pricing policies that force customers to buy large volumes of information actually worsen security and copyright protection problems because they create a strong economic incentive to cheat.

Intense interest generated by the Internet and the World Wide Web have brought pricing and related security issues to the fore. Expectations for a large, Internet-based market for information have excited not only futurists but also investors and information providers as well. Transacting business on the Web, while simple and cheap in theory, turns out to be technologically challenging and much costlier than many companies have bargained for (Ref 1). On the other hand, the Internet's origin as a free communications medium for researchers has contributed to an expectation among consumers for free or very low-priced information. Online services are responding to this demand by providing access to information available elsewhere at high prices as part of their basic subscription charges. An example is America Online's recent addition of historical stock quotes to its basic service (Ref 2).

The large amount of data stored on the current CD-ROM format, approximately 700 Mb without compression, also creates pricing dilemmas for providers. Typically, subscription pricing covers unlimited access to all the databases on a CD-ROM. Placing as much data as possible on a CD offers economies for providers and convenience for customers. And yet more data dictates higher prices, which limit sales. This problem will be greatly exacerbated by the advent of the next generation of this technology, DVD or digital versatile disc. DVD-ROM will increase the amount of data stored fourteenfold (Ref 3). How will pricing policies keep pace?

Linking price and use is the value customers place on information resources. Because information is increasingly disconnected from a physical medium — contrast a leather-bound book with a CD-ROM, for example — providers must constantly look for ways to support the price of a database with a high perception of value. The cost of subscribing to a single information resource for just one year can easily outstrip the cost of the hardware needed to access it. With hardware and operating costs dropping and performance rising, information providers can expect their customers to focus more attention on high fixed subscription costs.

A possible answer to the evolving pricing needs of the various distribution media lies in new commerce-enabling technologies such as information metering. By keeping track of usage by customers for billing purposes, metering facilitates the sale of electronic information. When deployed with supporting encryption, communications and database management technologies, metering makes it possible to sell information on a per-use or 'pay-as-you-go,' basis. This approach, which is only now becoming technologically feasible for CD-ROM and Internet-based delivery, offers the potential to bring into harmony the way information is used, valued and priced.

4. Fixed vs. per-use, or usage-based, pricing

The two most common ways of pricing goods and services are on either a fixed or a usage basis, and examples of each method abound in our everyday experience. Amusement parks, for example, may offer 'pay-one-price' entrance fees which cover the use of all rides, or they may offer a 'pay-as-you-go' approach and sell tickets for individual rides. In deciding between fixed and transaction pricing, assuming a choice is available, customers must consider how much use they are likely to make of the service. An expectation of heavy use points one toward the option with a price cap — fixed pricing — whereas an expectation of low use or uncertainty over the amount of use leads to the choice of variable usage pricing.

Fixed pricing has been used successfully for print media, such as books and periodical subscriptions, which typically do not lend themselves to usage-based pricing. Fixed annual subscription fees have been employed for

most information offerings on CD-ROM and for many early Internet offerings as well. In these cases, however, the lack of a practical method to implement usage-based pricing has dictated the use of fixed pricing rather than the needs of the marketplace.

Usage pricing in one form or another has been used for more than twenty years by the business, professional and financial sectors of the online industry. Extending the model used by long-distance carriers, online services have traditionally sold high-value information using a variety of connect time and volume charges, added to basic subscription or service fees. More recently, many leading online services have replaced time-based measures with output-based ones that charge for documents actually retrieved (Ref 4). The result of this shift allows search costs to be based more on qualitative results than on time spent searching databases.

The primary feature of fixed pricing — its predictability — has helped to position the CD-ROM as a cost-effective alternative to online services that base charges on connect time; thus, it appears to be market-sensitive (Ref 5). Fixed pricing by definition limits pricing flexibility, however, which affects both providers and their customers adversely. Information providers, for example, must pick one price to cover the use of large amounts of data over time. Too low a price loses revenues; too high a price loses customers. The conundrum for providers is that any single price will be both too high for some prospective customers and lower than what others, especially high-volume users, might pay. Information consumers, in turn, are forced to make all-or-nothing purchasing decisions. If they anticipate using a resource extensively and if they can afford the fixed purchase price, the decision will be positive: if these conditions are not met, their purchasing potential will remain untapped. Because fixed pricing limits revenues for providers and access to information for customers, its widespread use necessarily retards the market's growth.

Do markets prefer fixed or usage-based pricing? This question is not particularly apt since the two approaches are not necessarily in opposition. The key issue for provider and customer alike is how best to achieve a broader range of predictable pricing choices. Pricing schemes that combine fixed and usage-based pricing elements offer the best results for providers and a satisfactory range of choices for consumers. Such schemes are now common among the online services that charge annual subscription fees, which guarantee a base of revenues that can offset start-up and overhead costs, along with scheduled fees for documents actually retrieved.

5. Discounts for high-volume customers

Fixed subscription pricing discounts the cost of information to high-volume users in two ways. First, the fixed price functions as a price cap which means that the heaviest user of a service or resource pays the same price as the average or even occasional user. In this case the 'win' for the high-volume user comes at the expense of the provider, who would be able to charge more for a higher amount of information under a usage-based pricing scheme. Secondly, and more importantly, fixed pricing results in excessive volume discounts when applied to the pricing of information for network use.

Pricing schemes that list a range of prices for individual and multiple users give the impression of flexibility: one user pays less than five, five pay less than ten and so on. In fact, prices for network use of information are just as rigidly fixed as those for standalone users and arguably more so. Five heavy users of a resource, for example, pay the same price as five occasional users, assuming that pricing is based on the number of simultaneous users.

Because of the growth of networking, sales to networked users represent an increasingly important source of potential revenue; yet, providers commonly discount prices for large network installations by 75% or more, on a per-user basis, compared to subscription prices for individual users. It is imperative for providers to develop more effective, more logical and fairer pricing policies for information used on networks (Ref 6). Current approaches, which base pricing on a confusing array of criteria — such as the number of concurrent users, workstations, sites or employees in an organisation — are ill-suited to the needs of a growing marketplace and harm the interests of both provider and customer.

The thought of excessive discounts applied to high-value databases may please customers but the only true winners are the few organisations that can afford multiple-user licenses for all the information resources they desire. In most cases, smaller groups (even in large organisations) pay higher effective prices than larger ones and the relatively high fixed price of information limits the number of resources that can be purchased. A logical reaction to concurrent-user pricing has been the formation of consortia for the purchase of network licenses. These buying groups, mainly composed of academic libraries, take advantage of the high discounts that providers have extended to volume users. The possibility of diminishing revenues will undoubtedly force providers to re-examine their policies and search for better pricing approaches.

Usage-based pricing applied to the sale of networked information offers a way through the thicket. As long as information usage can be measured rationally and accurately, the intrinsic fairness of usage-based pricing will appeal to provider and customer alike. Customers who use a networked information resource intensively will pay more than the same number of customers who use it less. Usage-based pricing can eliminate the anomalies that occur when a large number of occasional users pay significantly more than a small number of heavy users, or when nodes are added to a network — requiring higher payment — but actual use of the information resource happens to decrease. Provided that all use is tracked for billing purposes, the need to police the number of authorised users or workstations disappears.

Usage-based pricing for network installations will also stimulate the growth of the market by providing

adequate financial incentives for providers and increased access to information for customers. It can also reduce the temptation to flout limitations on the number of authorised users, provided a secure metering system is in place. It is important to note that discounts to high-volume users are still feasible and will almost certainly be offered under usage-based pricing schemes, though the size of discounts need not be excessive or arbitrarily set. Other industries offer instructive examples. Carriers of long-distance telephone calls, for instance, offer a wide range of rate options and volume discounts for businesses within a usage-based pricing model that is universally accepted.

6. Examples of new pricing models

This section surveys three innovative information products or services that utilise usage-based pricing schemes.

6.1. Dow Jones News/Retrieval (DJNR)

Second only to LEXIS/NEXIS in the US market for online business and professional information, DJNR serves 233,000 subscribers. Its pricing has evolved since it abandoned traditional connect time measures three years ago in favour of an output-based scheme. With the Windows version that began shipping in June 1996, DJNR has completed the transition to pricing that relies primarily on charging customers for documents retrieved, an example of usage-based pricing. DJNR does include in its scheme elements of fixed pricing by charging customers a one-time start-up fee of \$29.95 and an annual subscription fee of \$19.95. The usage-based fees range from \$.15 for stock quotes to \$2.95 for articles found in DJNR's Publications Library (Ref 7).

DJNR continues the practice of offering fixed subscription pricing for large organisations. Rates start at \$500 monthly (\$6000 per annum) for five users. This tactic is an example of employing a price cap (actually a series of price caps) in order to provide a fair and logical discount to high-volume users. Because five low-volume users within an organisation could subscribe to the service individually at an initial annual cost of just \$150 and then pay for documents retrieved, customers are offered a choice of payment methods that correspond to their needs.

6.2. Harvard Business School Publishing

This subsidiary of Harvard Business School has begun distributing a CD-ROM catalogue containing *Harvard Business Review* articles and Harvard Business School case studies, and selling them on a pay-per-use basis. A major benefit for users is the ability to have immediate access to the full text of HBR articles and HBS case studies. Until now, customers desiring reprints would have to order the items by phone or mail and wait until they were delivered. Using technology provided by CD-MAX Inc., users can access these important materials on their desktops whenever they need them. Customers can search the index to identify items of interest and use a hyperlink function to connect to the full-text items.

Harvard Business School Publishing is using CD-MAX technology and supporting services to register and bill customers, track and report full-text usage and ensure data security. A special feature facilitates internal billing by enabling customers to track the cost of full-text usage by department.

6.3. IBM InfoMarket

IBM offers one of the first usage-based information services that is available exclusively via the Internet. Its InfoMarket offers information from more than 7000 sources, including such well-known North American providers as American Business Information, Hoover's and Engineering Information. Customers register with the InfoMarket service by providing billing information, including a credit card number. Like the competing service from NLightN, IBM uses a purely usage-based pricing scheme; no start-up or annual subscription fees are currently charged. Content owners or rights holders set the per-document prices, which range from gratis to \$3.50 and higher for proprietary information.

IBM InfoMarket has positioned itself as an information aggregator, providing a new distribution channel for companies that may not wish to invest the resources needed to engage in pay-per-use commerce on the Internet or a supplementary channel for companies that have their own Internet initiatives. The key to selling information on the open Internet lies in IBM's use of 'Cryptolopes,' a proprietary encryption method that seals content in envelopes that can only be opened by registered customers. This approach distinguishes IBM InfoMarket from competitors such as NLightN which do not currently prevent copying or retransmission of copyrighted materials.

7. Conclusion

New technologies create the need for as well as the ability for developing new pricing models. Some of the new approaches to pricing, emphasising a usage-based metric providing qualitative results, give information providers an effective means for increasing their market penetration and offer customers an affordable choice of options for the information they need.

David B. Boelio
CD-MAX, Inc.
219 South Street, Suite 101
Murray Hill
NJ 07040
USA
Tel: +1 (908) 464 1878
Fax: +1 (908) 771 0444
E-mail: Dboelio@aol.com or dboelio@cdmax.com

References

- [1] Fletcher, L. (1996) The new economics of online, *Searcher*, **4**(5), 30-44.
- [2] Fleming, M. (1996) Industry watch — end users: cannibals or new revenue? *Online*, **20**(1), 69-70.
- [3] Bell, A. (1996) Next generation compact discs, *Scientific American*, **275**(1), 42-46.
- [4] O'Leary, M. (1996) New names, new products, new pricing, and new looks for online business information, *Computers in Libraries*, **16**(6), 26-29.
- [5] Wiedemer, J. and D. Boelio (1994) CD-ROM pricing: the pay-as-you-go option, *CD-ROM Professional*, **7**(5), 45-52.
- [6] Rowley, J. and D. Butcher (1996) Pricing strategies for business information on CD-ROM, *Journal of Information Science*, **22**(1), 39-46.
- [7] SIMBA Information, Inc. (1996) DJN/R about to take wraps off a new and improved interface, *Electronic Information Report*, **17**(18), 1-3.



Information infrastructure: organisational capability for online information provision

Jane E. Klobas

The Graduate School of Management, The University of Western Australia, Australia

Abstract: *The World Wide Web (WWW) enables all organisations to become online information providers. For many organisations, the information provided will come not from external, published sources (the traditional source of online information) but from internal sources. Information formerly provided to external customers and clients (and to prospective customers and clients) during face to face meetings, from telephone information services, and through printed reports, brochures and publications, can now be provided online and made available locally and internationally on the Internet. Internal information formerly provided to internal staff in print form can now be provided online using WWW and related technologies to develop an intranet. Use of online media for provision of information from internal sources has repercussions for the organisation's information infrastructure. Successful online provision of internal information requires organisational, information technology and information management capability. From an organisational point of view, the key issues to be resolved concern organisational structure and information management rather than the technology itself. Online provision of internal information has the potential to affect the information flow massively, and therefore the structure and operation of an organisation. This paper identifies aspects of information flow that are affected by online provision of internal information, and recommends strategies for building organisational capability for online information provision.*

Keywords: Information infrastructure, intranets, Internet, World Wide Web, information provision, organisational capability

1. Online provision of internal information

The Internet is becoming widely accepted by businesses and government agencies as a tool for communication with clients, customers and members of staff (Refs 1, 2). The World Wide Web (WWW) was established to provide a method of sharing information in many formats between colleagues in many locations (Ref 3). Its success has led to its rapid growth as a medium for transfer of commercial and research information. The WWW and related technologies enable all organisations to become online information providers. Information once provided to external customers and clients (and to prospective customers and clients) during face to face meetings, from telephone information services and through printed reports, brochures and publications, can now be provided online and made available locally and internationally on the Internet. The opportunities for external information provision using the networks range from electronic publication of public relations and educational material to service provision.

Internet protocol technology can also be used to support electronic mail, distribution of policies and procedures, and file sharing within an organisation. Systems that use Internet protocols for internal information provision are called 'intranets' (Ref 4). Internal information formerly provided to internal staff in print form can now be provided online using WWW or groupware technology such as Lotus Notes to develop an intranet (Refs 4-6). Within organisations, price lists, policy and procedure manuals, acts with by-laws and other documents that are frequently updated and widely distributed can be made available electronically for people to consult as needed. For example, an electronic master copy of a policy manual can be maintained in an area accessible to all members of an organisation's staff connected to the network. Instead of printing updates or new copies of the manual each time a change is made, the master is updated with each change as it is approved. Each time members of staff consult the electronic master, they — and the organisation — can be assured that they are using the current procedure. Updates are more frequent and the chances of using outdated information are significantly reduced. Only one copy need be maintained (with appropriate backup). The cost of printing and distribution is reduced significantly.

Researchers who have looked at how the Internet is used to gather information at work have found that it quickly becomes part of day-to-day working life for people who have access to it. Internet information resources

are incorporated into work as easily as other working tools such as the telephone, the computer and the library (Refs 7-9). Use of the Internet and related technologies for information gathering can also fit readily into current organisational structures and practices. It is not so easy, however, for Internet information provision to be incorporated into an individual's day-to-day work or organisational structures and practices.

This paper identifies key components of organisational capability for online provision of internal information. Reference is made to four case studies of Western Australian government agencies' steps toward becoming Internet information providers, and to the lessons learned from their experiences. The paper concludes with recommended strategies for managers to prepare their organisations to become online providers of internal information.

2. Information infrastructure

'Information infrastructure' is a term used to describe the underlying information technology and telecommunications services that permit transfer of data between computers. The transferred data may include text, video, sound and other digitised media. Use of the term 'information infrastructure' to describe information technology and telecommunications services does not ensure that information is part of the infrastructure. Information infrastructure is the infrastructure that permits information to be transferred effectively. It therefore includes not just information and telecommunications technologies, but services, and organisational structures and processes that support effective information transfer.

Most discussion of information infrastructure has centred on the global information infrastructure of the 'information superhighway', of which the Internet is part (see, for example, Ref 10). National and local information infrastructures that are part of the Internet are also part of the global information infrastructure. The term 'global information infrastructure' will be used in this paper to describe the information infrastructure that permits an organisation to provide information to people outside the organisation, and to transfer information among people within the organisation but in different geographical locations.

The global information infrastructure is powerful because it permits information to be transferred between remote locations, without the need for synchronisation of the sending and receipt of information. The time and place matrix (Figure 1) illustrates the advantages of communication using the global information infrastructure. People can communicate with each other regardless of their physical location. Their communication can occur in real time or can be delayed to meet business requirements, differences in time zones, differences in work habits or simply personal preference.

While the global information infrastructure is essential if an organisation is to communicate effectively with people and organisations outside its boundaries, the organisation itself needs an information infrastructure if it is to be a successful online provider of internal information. The remainder of this paper considers organisational capability for online provision of internal information.

		TIME	
		Same	Different
PLACE	Same	Meeting facilitation	Shift work
	Different	Tele-conferencing	Electronic mail

Figure 1: Time and place matrix (adapted from Ref 11).

3. Organisational information infrastructure case studies

During 1995, a team of researchers from the Graduate School of Management at the University of Western Australia studied four Australian government agencies as they planned, developed or refined their presence on the Internet as online information providers. These agencies were pioneers in Internet information provision, and lessons learned from them gave us an insight into the organisational capabilities required for successful online provision of internal information. The agencies took a variety of approaches to Internet information provision, as

the following profiles indicate. The profiles, summarised in Table 1, describe each agency and its Internet presence at the time of the case study. Despite differences in their mission, size, structure and approach to Internet information provision, common lessons can be drawn from their experiences.

The Small Business Agency (SBA) works with new and existing small businesses to develop small business throughout the state. The agency has 40 staff. The agency's advisors provide information to people who run or who plan to set up small businesses in metropolitan and country locations. They provide advice by telephone, at in-person interviews, and through brochures distributed at trade shows and at agencies throughout the state. The SBA Internet presence was established by the agency's information manager who saw an opportunity to distribute information inexpensively to the agency's geographically distributed clientele. The information manager established the agency's presence by taking extracts from existing print publications, and modifying them so that they contained attractive navigational graphics and could be read on the World Wide Web. The Internet site was initially mounted on a PC within the agency's offices and connected to the Internet by modem.

The Commerce Development Agency (CDA) works with businesses to develop and enhance national and international trade relations. The CDA has 210 staff, located in Australia and other countries. The agency provides information to potential investors in the state as well as to businesses that operate within the state seeking to improve their international trade. Information is provided through formal training, brochures, introductions to approved advisors, and in telephone and face to face meetings. The agency represents the state at national and international trade fairs, and distributes promotional documents at these events. The CDA's Internet presence was established by the agency's public relations manager who saw the Internet as an opportunity to promote the state internationally as technologically advanced. The first material on the Internet site was adapted from promotional material developed for a significant international technology exhibition. The CDA Internet site was initially mounted on a server connected to the agency's Novell Network local area network.

The Resource Management Agency manages use of the state's natural resources. This is a large agency, with 800 staff in five major divisions. The agency has had several Internet champions. At the time of the case study, the technical information services manager, who had successfully established an Internet site for another large government agency, was planning to launch an extended Internet site. The new site would provide access to information about the agency and the state's natural resource. In addition, it would permit registered users to interrogate the agency's databases of public information about resource use and about issued licenses and permits. The site was technically advanced, with information accessed from several internal systems served through a Sun server, with the security shield of an additional Sun server that acted as a firewall to prevent unauthorised access. The information on the site came from existing agency print publications and online databases.

We included a fourth agency in the project. This agency, the Rates Valuation Agency (RVA), had not developed an Internet presence, but the agency's information technology manager had identified an opportunity for online information provision to the many organisations and individuals that sought information about the rated values of the state's land. The RVA has 180 staff, technologically-aware management, and a well-integrated information management infrastructure. This agency's systematic approach to establishment of an Internet presence provided a comparator for the more entrepreneurial approaches taken by the other agencies.

4. Organisational capability for online information provision

The GSM team identified three key elements of organisational information infrastructure or capability for Internet information provision (Figure 2): organisational structure and processes; information technology infrastructure; and information management infrastructure. The three elements of capability are interdependent. Furthermore, the conditions for establishing an Internet presence successfully are likely to be quite different from the conditions for successfully operating and maintaining that presence.

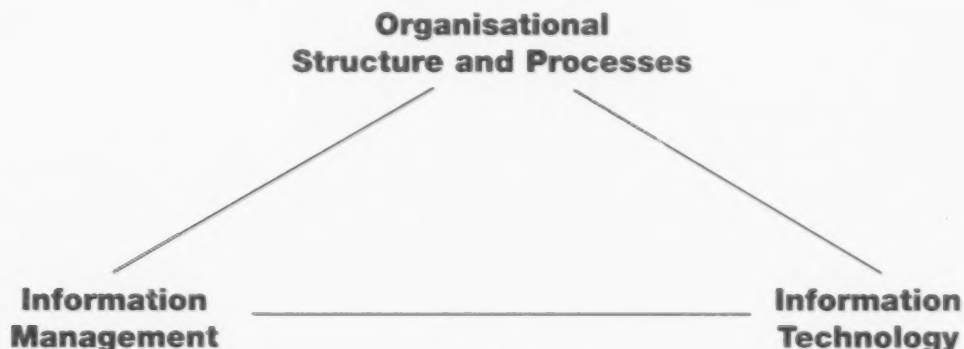


Figure 2: Capability for successful information provision using computer-based communication networks.

Table 1: Summary case agency information.

Agency (a)	Mission	Size (b)	Technology for Internet information provision	Internet information provision project leader	Information provided online from the Internet
Small Business Agency (SBA)	develop small business throughout the state	40	PC and modem	information manager (librarian/database administrator/network administrator)	information about the agency and its services(extracts from existing print materials)
Commerce Development Agency (CDA)	support the state's business organisations as they develop and enhance national and international trade relations	210	modem connected to office local area network (Novell Netware)	public relations manager	information about the agency, and copies of promotional material for the state (from existing print publications, with supplementary introductions for the Internet site)
Resource Management Agency (RMA)	license, monitor, and manage use of the state's natural resources	800	Sun server with separate firewall	technical information services manager	information about the agency (from existing print publications, with supplementary introductions for the Internet site)
Rates Valuation Agency (RVA)	value land and properties for rating and taxation	180	none	information technology manager	none

Notes: (a) agency pseudonym, (b) approximate full-time equivalent staff

4.1. Establishing an Internet presence

Internet-based information technologies such as the World Wide Web have three key advantages that make initial establishment of an Internet presence attractive to, and possible for, individual entrepreneurial members of an organisation's staff. Internet technologies are scaleable, they have a low entry cost, and concerns about the technology itself tend not to surface during establishment.

It is possible to establish an Internet presence with a PC, a modem, an account with an Internet service provider and a telephone or other telecommunications link. Quite serviceable software is available free of charge from the Internet. For a commercial organisation in Australia, the direct operating costs of this connection are less than A\$1500 a year. The Small Business Agency established its Internet presence with a technical set up just like this. It is equally possible to establish an Internet connection from most large organisational computer systems.

Organisations that have existing printed information, particularly if that information already exists in a digital form, can transfer the information with a little effort into simple text form or hypertext mark-up language (HTML) that can be read with any World Wide Web browser. Many Internet information providers have established their online presence by transferring existing published or written documents to their Internet server. This is a common approach for organisations that have well established mission statements, descriptions of activities, and policy or educational documents. This approach has the advantage that the information provided is readily available, and already approved for distribution to the public. If it is out of print or costly to distribute, it can be made available to people who would otherwise not have access to it.

The ease with which the initial set of information can be made available online through Internet technologies is not dampened by major technological concerns. Concerns about security have receded as information technology managers have learnt to construct effective firewalls to screen unauthorised access, and to use effective encryption technologies (Ref 12).

Many organisations that have established an Internet presence have an internal champion or 'intrapreneur', who has access to the technology and the content, and the ability to convince decision makers to invest in a pilot site. The Internet intrapreneurs we met during our research were more concerned with establishing an Internet presence than with maintaining the information for longer-term use.

4.2. Maintaining a successful network presence

Successful online provision of internal information requires more than an intrapreneur with some existing technology and content. Successful online information provision makes available up-to-date and accurate information that is needed by members of the organisation, its clients and suppliers, and/or members of the public. The online presence should enhance the organisation's ability to fulfil its mission and to meet its objectives. It should meet current standards at lower cost, and enable the agency to exceed current standards without increasing costs. Online information and service provision should be fully integrated into the organisation's information and service provision functions. It should be supported by appropriate information technology, information management infrastructure, and organisational structures and processes, including people with the ability and skills to maintain secure, accurate and current information and information technology services.

At present, only those organisations that have been established specifically to provide online information and services approach these standards. Network information resources developed by existing organisations seldom meet them. You do not have to visit many Internet World Wide Web sites to find several that are under development or out of date — or both under development and out of date! Many organisations that have mounted existing print documents on the Internet have structured and presented information in a way that does not match the strengths or the limitations of computer screens and underlying indexing and searching technology. Problems include graphics that do not fit on a screen, contents buttons that do not appear on the first screen that the user sees, long documents through which users are expected to scroll without guidance, and colours that blend into the background so that some words or letters are unreadable.

From an organisational perspective, online provision of internal information takes advantage of a new communication medium. The global information infrastructure is a computer-based communication network that can be seen as a complement to other organisational communication media such as face to face meetings, television and radio broadcasts, print material, telephones and facsimiles machines. Just as decisions are made to incorporate each of the other media into the organisation's practices and programmes for information provision, so computer-based communication media can be included in these programmes. Computer-based communications should be integrated with other modes of communication. For this to happen, an organisation's staff must become as familiar and confident with computer-based communications media as they are with other media. This requires education and training in the medium as a communications medium rather than training in operation of the information technology.

To integrate network-based information and service provision into the day-to-day work of organisations, staff responsible for information and service provision must have physical access to the technology. A simplistic approach would be to enable existing desktop computers to access the Internet, or to provide staff without desktop computers with computers that provide Internet access. Without a clear programme for networked information and service provision, this approach would help with electronic mail communication and information gathering, but not with information provision. The time and place matrix (Figure 1) shows that the nature of requests for information and service, and the way in which they are fulfilled, can be altered dramatically if

networked computers are used to provide information effectively to people from many locations who seek information at any time of day.

For printed policy manuals to be replaced by electronic manuals, each member of staff must have convenient access to a terminal and skills to retrieve the information. Behind the scenes, staff preparing the manuals must have skills in electronic information management, ranging from design to version control and other records management skills. A mechanism will be needed to alert staff to changes to the electronic manual; they will no longer have the visual cue contained in a set of updates. Technical staff must be competent at maintaining a secure and reliable system.

While it is possible to fit electronic mail communication and use of networks for information gathering into organisational structures and practices without re-design, use of the networks for information provision has the potential to affect the entire structure and process operation of an organisation. Electronic networks provide an opportunity to improve services and organisational structures radically. Managers who are aware of their organisation's strategy, the capabilities of the medium, the needs of their clients and the needs of shareholders, government or other stakeholders may find a solution to long running problems, provided they have the courage to engineer radical change.

Even if radical change is not called for, online information provision demands changes to information processes. It is not enough for an intrapreneur to gather and mount existing information on a network server: the information must be kept up to date and accurate. It must present an accurate and positive view of the organisation to the community. The information should be presented in a way that suits the technology and that enables people to obtain information with little difficulty. This requires networked information provision to be integrated into the organisation's information provision functions and therefore its day-to-day information flow. Where networked publication and distribution replaces print publication and distribution of information, staff formerly employed in print distribution roles will need to be retrained to meet the requirements of electronic publication, including layout and design, updates and version control. Staff in service and information provision roles will need to know what information is on the network, how to obtain it, and how to help others to obtain it.

The agencies in our studies expect Internet information provision to reduce the need for over-the-counter information services. Staff who provide counter service now will have to be re-deployed, and possibly re-trained to maintain and provide information in electronic form. To maintain even an apparently small site will have staffing and organisational ramifications.

The information technology and information infrastructure must support changes in use and demand. Information technology staff must be capable of maintaining secure access during the hours the service is provided — possibly 24 hours a day, seven days a week. They must be able to meet demand for access while at the same time maintaining appropriate levels of security. Records managers and information technology staff must work together to ensure that different versions of documents and information are appropriately controlled, and that appropriate records of electronic transactions and correspondence are kept.

5. Summary and conclusion

Online provision of organisational information, using computer-based communication technologies such as the Internet (or new technologies that develop to take advantage of the global information infrastructure), will become as commonplace as other forms of organisational communication. Even if they do not establish themselves as online information providers in the short term, organisations should begin to develop the capability to do so now.

Successful online information provision requires organisational, information technology and information management capability. From an organisational point of view, the key issues to be resolved concern organisational structure and information management rather than the technology itself.

Managers who want to guide their agencies through the period of change that will accompany development of online provision of internal information will need to develop strategies for personal and organisational development. We recommend the following strategies:

- (a) *Develop in managers an understanding of the new information and business environment.* Participate in education and training in the role that global and organisational information infrastructures play as communication media, in business process re-engineering, and in change management.
- (b) *Develop managerial, professional and operational staff understanding of the organisational role of computer-based communication media.* All organisational personnel will need a new understanding of the role of computer-based communication media and their relationship to other forms of communication currently in use. Education should focus on how computer-based communication media can be used in organisational communication, rather than training in operation of the hardware and software that form the information technology infrastructure for computer-mediated communication.
- (c) *Re-engineer business processes whose end-product is information provision.* Re-engineer organisational processes to incorporate online information provision in day-to-day work processes. In most organisations, information products are the end-product of processes that involve revision and reformatting of information that originates early in the process. Online information provision permits information to be provided to the organisation's staff, clients, suppliers, stakeholders and others as it becomes available, without the significant further processing currently required to deliver information from given locations staffed during set hours, or in the form of a printed document with an acceptable shelf life.

- (d) *Develop policies and procedures for management of online information about the organisation.* Information management policies may be based on organisational policies for issue and version control, information security and record retention, where they exist. If the organisation does not have such policies, they will be needed in an electronic environment where the content of an information resource can be changed much more quickly and easily than that of printed resources.
- (e) *Provide appropriate staff with suitable information technology and skills for online information access and provision.* The staff who provide online information form an important component of the organisation's information infrastructure. They will need access to suitable technology for online information, and instruction in its use. They will also need training in information management, which includes implementation of the organisation's information management policies and procedures, including how to maintain the quality of online information.

Jane E. Klobas
 The Graduate School of Management
 The University of Western Australia
 Nedlands 6907
 Western Australia
 Australia
 Tel: +61 9 380 3567
 Fax: +61 9 380 1072
 E-mail: jklobas@ecel.uwa.edu.au

Acknowledgement

The case study research was supported by the University of Western Australian and the Australian Research Council. Thanks to Kate McGoldrick and Greg Jarosch, research assistants at the Graduate School of Management, the University of Western Australia, who are undertook the case studies. This paper is based on a presentation to the Royal Institute of Public Administrators, Australia (WA Branch).

References

- [1] Cronin, M.J. (1995) *Doing More Business on the Internet*, Van Nostrand Reinhold, New York.
- [2] Morris, P. (1995) WA business slow to realise net benefits, *The West Australian*, 3 October, p. 60.
- [3] Berners-Lee, T., R. Cailliau, H.F. Nielsen and A. Secret (1994) The World-Wide Web, *Communications of the ACM*, **37**(8), 76-82.
- [4] Cortese, A. (1996) Here comes the intranet, *Business Week*, 26 February.
- [5] Spurr, K., P. Layzell, L. Jennison and N. Richards (Eds.) (1994) *Computer Support for Co-operative Work*, John Wiley & Sons, Chichester.
- [6] Ellis, C.A., S.J. Gibbs and G.L. Rein (1991) Groupware: some issues and experiences, *Communications of the ACM*, **34**(1), 38-58.
- [7] Cronin, B., K. Overfelt, K. Fouchereaux and others (1994) The Internet and competitive intelligence: a survey of current practice, *International Journal of Information Management*, **14**, 204-222.
- [8] Klobas, J.E. (1994) Networked information resources: electronic opportunities for users and librarians, *Information Technology and People*, **7**(3), 4-17.
- [9] Sproull, L. and S. Kiesler (1991) *Connections: New Ways of Working in the Networked Organisation*, MIT Press, Cambridge, MA.
- [10] Global Information Infrastructure Commission (1996) *Global Information Infrastructure Commission Home Page*, <http://www.gii.org/>, page current at 25 June 1996 was last updated 26 February 1996
- [11] DeSanctis, G. and B. Gallupe (1987) A foundation for the study of group decision support systems, *Management Science*, **33**(5), 589-609
- [12] Liu, C., J. Peek, R. Jones, B. Buus and A. Nye (1994) *Managing Internet Information Services*, O'Reilly and Associates, Sebastopol, CA.



Marketing principles as applied to the corporate information center

Suzan A. Brown

Director, Marketing, Chemical Abstracts Service, USA

Abstract: Marketing principles, as applied by major businesses around the world, can also be used by information professionals to grow and expand their presence within their own organisation. Like the corporation itself, the information center must be concerned with budgets, production and satisfying 'customers'. It has also been true, since the beginning of corporate 'libraries', that information professionals have recognised a need to market their services. Marketing techniques that have worked for industry can also be employed successfully by information centres in reaching and influencing their in-house clientele. The paper will focus on parallels between marketing in the industrial/research arena, and the needs of information professionals to expand business from existing information center customers as well as expand into new markets in the corporation. Market research, product concept testing and customer surveys will be discussed as proven marketing techniques. The importance of managing the entire information center product line and techniques for promoting key services of the information center will be reviewed.

Keywords: information center, information professionals, marketing, promotion

1. Introduction

Like it or not, marketing and all the activities that go with it are a prominent part of everyone's life these days. While you may not be a marketing professional yourself, you probably work for an organisation that markets a product or service. If nothing else, you are certainly exposed to someone else's marketing, most visibly in the form of advertising that confronts you at every turn. How many thousands of those messages you are exposed to during a typical weekday, no one can say for certain — but you see my point. What many information professionals may not realise, however, is that they themselves are engaged in marketing every day, and it should be a valuable tool in an enterprise so customer-dependent as a library or corporate information center. In light of that reality, it is important to recognise the marketing side of your work, embrace it, and use it to your best advantage.

2. Marketing and the information center

First, let's think about the word 'marketing', because it is a term you hear so often you may take the meaning for granted. Most people think of marketing as *advertising* or *selling* and, since both of those activities can be irritating when you are at the receiving end, marketing may have distasteful connotations for some people. In reality, marketing is much more than selling. It is an exchange of value in which both parties gain something, and virtually any organisation — even a university or a government or a church — is involved in offering value to a group of constituents. The difference between selling and marketing, as explained by Theodore Levitt in *The Marketing Imagination*, is that 'selling focuses on the needs of the seller, marketing on the needs of the buyer.' In that sense, Levitt goes on to say, marketing is concerned with 'satisfying the needs of the customers by means of the product and the whole cluster of things associated with creating, delivering, and finally consuming it' (Ref 1). To market effectively, you must first of all know the business your organisation is in, so you can contribute to its success.

Certainly, libraries or information centres have always been concerned with satisfying the needs of their clientele and offering value in exchange for their support. In fact, a library exists for no other purpose — does an information center really exist if no one uses it? Well, not for long. Today more than ever, an information professional must be concerned with marketing his or her services within the organisation. I say 'more than ever' because there are more alternative sources of information than ever before. That means two things: (1) your clients need more help choosing among the alternatives, and (2) your information services face more competition. Those are classic marketing challenges faced by all kinds of organisations.

Your information center is a microcosm of the organisation to which it belongs. Like those who manage the organisation, you must be concerned with budgets, production, managing the product line, marketing and selling, and keeping your customers happy.

3. A sense of mission

Once you have come to recognise that your information center is a marketing enterprise, you begin to wonder how to do marketing. After all, this wasn't why you decided to study information science! How do you move toward being a marketer and not just the object of someone else's marketing? Largely, it is a matter of perspective and intention. Every good marketing plan begins with a mission statement, ideally an expression of purpose that applies to the entire organisation and is reflected in the plans of every organisational division, not just marketing. Some organisations don't have such a statement because they assume all their employees know what the corporate body is trying to accomplish. That assumption is a bit risky, however, because it is all too easy to lose site of the big picture in the clutter of details. If your company exists to produce cosmetics, then you must remember you are in the cosmetics business, not the information business *per se*. Your work is the means to an end. Think about the end result and use your company's mission statement as a guide to devise a mission statement specific to your own operation.

A mission statement should be brief — one sentence that asserts what you intend to accomplish. To paraphrase Walt Disney, the mission of Disneyland at one point was expressed as 'giving people such an enjoyable experience they will want to tell their friends about it and come back someday.' At Chemical Abstracts Service (CAS), to mention another example, our mission is to meet the needs of users of chemical and related scientific information world-wide. Possibly, your mission as an information specialist is similar, if you work for a company engaged in scientific research and development. I suggest you begin your marketing plans by crafting an information center mission statement that you can share with all your colleagues. Keep a few characteristics in mind as you write it: make your statement one sentence, express what you want to be, and acknowledge the people you are trying to serve. Obviously, the mission of your information center or department must be in sync with that of your organisation as a whole. If you or any of your staff don't know what that organisational mission is, find out right away!

Remember, it's only in terms of the organisation's overall mission that your own goals will be meaningful. Does your employer make pharmaceuticals, or agrochemicals, or does it teach students? Does it exist to improve health, or increase agricultural productivity, or educate scientists? How will your information center facilitate that work, improve the production process, reduce cost or otherwise contribute to the 'product'? Find out by talking to anyone in the organisation who can relate your work to their results. Then make it clear to everyone what you intend to contribute — that is your mission statement.

A good example of a corporate library mission statement was provided to us by the head of the Hewlett-Packard Research Library:

'to lead Research Services in the efficient and effective provision of information to HPLabs' employees in a manner that accelerates and fundamentally improves the research process' *Eugenie Prime, Manager of Corporate Libraries, Hewlett-Packard Company*

4. Defining your customers: market segmentation

In order to satisfy your customers, you must first identify them. This step is called market segmentation, and it has been described as follows by Hassan and Blackwell in *Global Marketing*:

'Segmentation is defined as the process of dividing the total market into subgroups for which a marketing mix that has special appeal to the most attractive (profitable) of the market segments can be developed' (Ref 2).

In other words, don't just say that you serve everyone in the company. Your clients are actually comprised of different subgroups or segments that have different needs. You may define many segments, and some are more important than others. Try to gauge their importance based on such criteria as the financial impact they represent, their strategic importance to your organisation, and their potential for innovation that you can help to realise, among other factors. The term 'marketing mix' refers to the set of activities that will influence these different groups. Different segments might include, for example:

- administrative management, concerned mainly with the organisation's overall productivity and viability; interested in activity of competitors, new business opportunities, regulatory requirements;
- R&D management, concerned with currency and completeness of scitech information; likely to have a special interest in patents;
- scientists and engineers, primarily interested in project-related information; narrowly focused on previous research in the same area but open to serendipity.

These are only some of the segments you might define. In some organisations, important groups would include the legal department, process engineering, quality control, the compliance officers, marketing, or you name it. In universities you would logically treat undergraduate students, graduate students and faculty as different segments. In all cases, segmentation depends on the interests you can identify that are common to a group and relate directly to your services. The act of identifying your customers necessarily includes assessing their needs. In short, it is their information needs that define them.

Online Information 96 Proceedings

Page 182

You can assess information needs one-on-one as you work with the customers that come to you. But what about the potential customers who never think about consulting the information center? For them, outreach may be necessary in the form of a consumer survey. This can be a simple questionnaire distributed in hardcopy, perhaps in your organisation's employee newsletter, or by e-mail. The questionnaire can be designed to assess the needs of both current and potential customers, and should address such issues as:

- satisfaction with current services offered;
- other services used (outside of your information center);
- perceived needs for other kinds of information or additional services;
- any way in which the information center can provide better service.

Interpreting the results of a survey deserves more discussion than can be addressed in this paper; however, it is worth noting some findings reported in a recent study by American Airlines. According to this study of airline customers:

- one unhappy customer tells 9 to 13 other customers about his or her bad experience;
- only four percent of unhappy customers complain to the company;
- for every person who complains, there are 24 other unhappy customers who don't say anything;
- 75-90% of unhappy customers who say nothing will never do business with the company again;
- 82-95% of those who complain will come back if their problem is resolved quickly (Ref 3).

Evidently, receiving even a few complaints can mean more dissatisfaction than you would like to believe! Dealing with the complaints you hear is crucial.

The act of distributing a survey will in itself draw attention to the existence of your service and encourage everyone to think about its potential for them. The survey should mention some of your information 'products', which we will now examine.

5. Defining your products

Considering the segments I've mentioned, among many others you might think of, one can envision different kinds of information products suitable to each group. Any of them may require a *comprehensive search* from time to time, a task that end-users are not well equipped to handle for themselves. This is especially true since online files have changed in some respects over time, for example indexing practices for a given database may have been different in the past. Moreover, in the case of Chemical Abstracts for example, an exhaustive search extending back beyond the mid 1960s would require searching printed indexes, too. Chemists who may perform simple searches online on their own quite effectively may need your help using a printed subject index from 1960.

You may find that certain clients could benefit from an *ongoing current awareness service* tailored to their continuing interests. This is especially likely for anyone involved in R & D, where breakthroughs are always in the making. For the project-oriented scientists, who are interested in new ideas as well as specific data, a *table of contents service* is a promising option. They may be amazed to discover that the advantages they once enjoyed browsing their favourite journals — for which their subscription fell victim to the cost-cutting axe — are now available in an online format. The point is, you are engaged in *service marketing*. You can help your customers not simply by giving them the information they have requested on a given day, but also by identifying services that will benefit them continually.

Consider how your mix of services might be revised. What functions do you do well? What else can you do easily? How do these compare with the key customer needs you've discovered?

Think about other kinds of service marketing you've encountered outside of work. Insurance agents, airlines, restaurants, even good auto repair shops go beyond serving your immediate request and look into your ongoing needs. Certainly that is a form of selling, but it means establishing a *customer relationship* that works to mutual advantage. What is it that distinguishes one bank from another? All offer the same basic services. You notice the difference mainly when something goes wrong — handling problems efficiently and gracefully is a big part of marketing any service.

Once you've begun to explore your clients' needs in more depth, you'll also begin to think of other products or services you can offer. Learn from what your colleagues are doing in other organisations, and use your imagination. Look to your competitors, too, and those might be independent information brokers. What can you offer that others might not do as well? You may even find that printed bulletins are still an excellent medium for some purposes! Among other things, they are good for summarising competitive information for management.

Just as any company would, you should test a major new product before launching it. Don't announce an 'answers on demand' capability to the company at large until you've tried it out with a small group of customers; otherwise, you may discover the demand is more than you can handle. Other factors to test for are cost, customer satisfaction, and the impact on other products or services you provide. Introducing a new alternative may lead to the demise of an existing product. That result may be acceptable, but you need to plan for it.

It may become necessary at some point to discontinue a service, and in terms of increasing your efficiency that can be just as important as creating one. When this becomes necessary in the course of product management, it presents a challenging customer relations problem because people tend to hold on to what they've become accustomed to receiving. Reasons to discontinue a service might involve several factors such

as impact on your costs, relationship to your mission, the 'fit' with other services you perform, or the importance of the groups or individuals using the service. 'Importance' is relative and subjective, but it's necessary to know who considers a service important and why.

What if the time consuming, printed clipping service you've been providing a certain manager is no longer possible because you've had to cut staff? Even a service that is seldom used seems to become essential once the customer realises you're going to discontinue it! While you explain the necessity of dropping the old service, offer some alternatives at the same time. Perhaps customising an automated electronic update service would provide an acceptable substitute.

Speaking of alternatives, I touched on competitive analysis before but it deserves emphasis. Take the trouble to know all about the alternatives to your own information service that your customers could conceivably choose. As an information professional, you need this type of analysis for two reasons: (1) you may want to incorporate some of those services into your own operation or (2) you may need to explain why your current offering is better. I suggest you prepare a table listing competitive services along one axis and features and benefits along the other, remembering that cost is an important benefit (or disadvantage). Competitors could mean online services you don't use for some reason, outside information brokers, and any alternative information source your customers might possibly use. That would include subscriptions to printed publications, primary or secondary. If you're asked why one alerting service is worth a dozen journal subscriptions, be prepared to answer with actual figures at the ready.

Deciding what services to offer, and how to offer them, depends a great deal on the values you wish to uphold: that is, the qualities you consider essential to your operation. You may decide that all your services will be delivered on time, cost-effectively, and with personal attention to the customer's satisfaction. If so, that commitment may affect the quantity and type of service you will offer.

To an increasing extent, the Internet is an important part of the information marketplace, and it relates directly to your sense of value. Some of your co-workers — and this may be especially true of scientists — probably believe the Internet is the ultimate solution to their information needs. They know they can submit questions to their colleagues on the Net and get quick answers about chemicals and many other topics of routine interest. Unlike you, they may not know the importance of indexing, document analysis and other value-added aspects of professionally built databases. And what about the reliability of the information sources, which may be far more important than currency or ease of access? Much of your role will always be education in these matters.

Don't think of the Internet as a threat, however, because the technology may result in new products that your information center can offer. If your organisation supports it, you should consider using the Intranet, an in-house Web site accessible only from within your organisation, that can distribute all kinds of information instantly with the advantages of electronic updating, graphic appeal, and ease of use we've come to associate with the World Wide Web. Imagine using such a tool to post news of interest throughout the company based on your company-specific monitoring of the competition. The Intranet is a new channel of communication worth a paper in itself, and it may become especially important to the corporate information center.

6. Objectives for marketing planning

Knowing your customers and their needs, you can proceed to set some concrete objectives for your information center. Your objectives will be unique to your situation but in any case objectives should be:

- specific (not a job description);
- results-oriented (not a list of activities);
- quantifiable and measurable;
- achievable within a definite time frame;
- fewer than 10 (and should fit on a single page).

Make them realistic and obtainable as well as challenging. Your mission statement devised earlier establishes the ultimate destination of your plan but you need to specify some quantifiable milestones to aim for on the way. The objectives are the core of a marketing plan that will lead you to success.

Consider some hypothetical objectives:

- *broaden customer base:*
 - conduct searches for 100 more customers than served in 1995;
 - begin serving a division we've never previously reached.
- *demonstrate value of information center:* document at least three information projects that resulted in \$100,000 or more in increased revenue or cost savings for the company; distribute report to key managers by (date);
- *improve productivity:* train 50 scientists to conduct their own searches, thus freeing the information center staff for more complex searches.

After you've devised your objectives, don't keep them to yourself. Sharing them with staff and management will help you accomplish your goals. Others may offer to support you in ways you didn't even expect. Also, having publicised your objectives, you will automatically feel more committed to achieving them.

7. Selling and promotion

Once you've identified target markets, defined new and existing customers, set some objectives, and have devised the best suite of services you can, you need to publicise what you have to offer. As I've tried to demonstrate in this discussion, selling and promotion are not the first steps in a marketing plan but are properly brought into play near the end. In fact, you may do a substantial amount of pre-selling during the earlier process of assessing your customers' needs. After that, the whole arsenal of consumer promotion can be put to work. Information specialists we've worked with have used every technique you can imagine, including:

- posters at the 'point of sale' (e.g. outside your door) or wherever employees gather;
- occasional flyers or electronic messages distributed broadly;
- advertisements or articles in the company newsletter;
- information fairs that exhibit your services to all interested staff: some information professionals invite representatives from their vendors to exhibit at these in-house events, because their services are part of yours;
- give-away items reminding people how to reach you'
- product sampling, such as a free current-awareness search or an alerting service for a limited time period;
- promotional packaging, including identification of your information center on the search results you deliver;
- presentations to selected groups within the company.

Concerning the content of promotional messages, remember that everyone loves news, even if they hate advertising. So make your message sound like news. Even a service you've been offering forever will be news to many in your organisation, and there's always something new to say about it.

If some of this sounds too much like hucksterism for an information scientist, remember what your competitors are doing! Outside sources of information, including information brokers and other commercial services your co-workers might use for themselves, are constantly promoting to your customers. But since you are located on the scene, you have an enormous insider advantage. Make the best of it! Start by finding out how your customers found out about the other services, if they are in fact using them.

You may find it beneficial to position your outreach initiatives as education and identify them as such in your organisation. Arrange 'seminars' for relevant groups of potential customers in which you explain the array of information resources available and their benefits for the audience. Your event won't be perceived as a sales pitch but will nevertheless win new customers. The seminar approach is widely used in service marketing. Many pharmaceutical companies, for example, use it to reach physicians very effectively.

8. Know when it works

What you do after all this marketing effort is just as important as the effort itself. Measurement is essential, so you find out what worked and whether a given project was worth your expense. Follow up with your customers and invite feedback whenever possible. Do this in a formal way that you can document. By that I mean not only asking customers what they thought of the search results you provided, but also the results of the results. Did an article or patent you found for a chemist lead him or her to a discovery in the lab? Or did it save a million in research and development by revealing that a certain chemical has already been patented or reported in the literature? Track the results carefully and then make sure everyone who counts knows about your success. This isn't just for job security, but will help your group to help your business. In the process, don't neglect thinking about efforts that fell short, because they can teach you as much as your triumphs. As any information provider will tell you, products that fail are very often those whose usefulness seemed obvious to the producer — so obvious the customers were never asked their opinion.

9. Conclusion

Think of your information center as a business in its own right with customers, a mission, competition, and a range of products and services to market. Use marketing techniques such as segmentation to identify your customers, set concrete goals for your marketing plan, and employ all available promotional and sales techniques within your organisation. Above all remember that an information professional is engaged in service marketing that depends on cultivating customer relationships. Solicit feedback, document your results, learn from what went wrong, and make your successes count toward the goals of your organisation.

Suzan A. Brown
Chemical Abstracts Service
2540 Olentangy River Road
Columbus
OH 43202
USA
Tel: +1 (614) 447 3600
Fax: +1 (614) 447 3709
E-mail: sbrown@cas.org

References

- [1] Levitt, T. (1986) *The Marketing Imagination*, The Free Press, New York, p. 153.
- [2] Hassan, S.S. and R.D. Blackwell (1994) *Global Marketing: Perspectives and Cases*, Harcourt Brace & Co., New York, p. 53.
- [3] Barnett, C. (1996) How to complain effectively, *OAG Frequent Flyer*, July, 22-29.



The intranet: a new concept for corporate information handling

José Barberá

Fundesco, Spain

Abstract: In the last year the World Wide Web model has evolved within companies from a repository for notice boards to a new tool that can improve work productivity. Intranets, the internal or corporate internets, are likely to be the key information technology revolution for the remainder of this century. The intranet concept is derived from the present Internet as a natural step in its own evolution. The same TCP/IP communication protocols and applications are used, in particular the Web server/client model. This paper gives first the rationale for setting up corporate intranets, providing the main reasons why the Internet technologies are having such a remarkable impact on the scope of business networking applications and why intranets are becoming so popular among a broad variety of companies. Secondly, the key distinctive functional features of intranets are presented. A discussion on typical intranet applications follows: the new functionality offered to business people and other non-typical IT users is stressed, and significant differences with existing groupware solutions are discussed. The paper ends up outlining a practical ongoing project to digitise documents (thus reducing the need for printed materials and allowing the inclusion of multimedia contents) which can be delivered to each desktop and accessed by all individuals of an organisation in a cost-effective, flexible and friendly way through an intranet Web interface.

Keywords: database, client/server, groupware, distributed information, multimedia information, Internet, Intranet, TCP/IP, WWW, World Wide Web

1. Rationale

The Internet boom, and especially the World Wide Web (WWW) fever, has spread out in the last two years among a variety of organisations of different size and scope. Today companies have learnt that a connection to the Internet means worldwide access to many information sources — often valuable information — which may be useful for the organisation. On the other hand, publishing information and keeping in touch with customers and suppliers through the Net are new possibilities which companies see now as additional marketing channels.

In that situation, many organisations ask themselves: if the Internet is based on an open, proven and reliable technology, if there is a broad range of applications and services, and if the WWW has become a *de facto* standard to access and disseminate information world wide, why not take advantage of all that potential in order to implement internal information procedures? The intranet concept springs up thus as the evident answer to that question. Put simply, 'intranet' is the descriptive term associated with internal corporate networks which are implemented using the Internet technology and services, in particular the WWW system, but adapted to the specific physical boundaries and internal procedures of each organisation, so that certain information is available only to its employees and not to the public Internet. Implementing internal publishing procedures and other collaborative tasks on the Web via so-called intranets is proving to be simpler and more efficient, at the same time allowing easier communication with third parties.

2. Functionality of the intranet

From a functional point of view an intranet allows the immense amount of informational resources that flow within an organisation to be transported and delivered to each individual's desktop with minimal time, cost and effort. Consider this scenario. A company has 20 sites and 3000 employees who need timely access to corporate information such as company news, policy changes, training manuals, organisational procedures and the like: even to simple but crucial documents such as telephone books, pricing information and product specifications. Individual workgroups and departments require secure and limited access to confidential data. Each branch, possibly in different geographical areas, very often has its own database containing reports and data that must be shared with the head office (Refs 1, 2).

So far the normal procedure is the hard copy production and physical distribution of printed material such as

handbooks, lists, guides, etc. The production of this printed material is not only expensive and time-consuming but also highly inefficient as it needs frequent updates. And how does one guarantee that all the offices receive the information on time, quite often before it becomes obsolete? How can one ensure that people know important policy details or other relevant information that have recently changed? The straightforward answer is that with the existing technology one cannot. Nevertheless it is a proven fact that information must be kept as accurate as possible and flow rapidly in a business world which is more and more dynamic and competitive.

The answer to those problems is the Intranet for it is a solution that (Ref 1):

- (a) allows the delivery of information upon demand, and when needed;
- (b) can guarantee that the information is accurate and up to date;
- (c) ensures that the information can be kept in a single source (although it is not necessary that source to be the only source of all information);
- (d) allows the information to be maintained by those individuals and groups that produced it originally.

Today those problems can be solved using the generic Internet technologies which are giving rise to drastic changes: a true revolution for the information systems of the business world within the further revolution of the Internet. An intranet lowers the cost and shortens the time of distributing information. Instead of producing hard copies and delivering them by ordinary mail, documents such as newsletters, minutes, policy procedures, projects and other collaborative tasks can be stored on internal Web pages and shared by people who will peruse them at the right moment as long as they have access to the corporate intranet.

In addition to that, the intranet allows the external information existing on the World Wide Web to be organised according to the company's needs. On the internal Web server there may be links to relevant online reports and publications existing in public Web servers: for instance information about competing organisations. In summary, the intranet is becoming so popular among companies in fields so diverse as publishing, finances, insurance, healthcare, software and so on because it allows to save costs and time, to centralise information, to share internal information and to organise the information via Web pages.

3. Distinctive features of the intranet (Ref 3)

One of the key aspects of the Internet is the use of the well-known TCP/IP protocol suite to communicate between heterogeneous computer systems connected to at least one of the almost 100 000 networks that make up this worldwide information infrastructure. A similar situation can be seen, in a smaller scale, within corporate organisations: there is a great diversity of hardware and software computer resources, usually interconnected via local area networks: diversity which is also reflected on the human users and their functions.

The challenge of the Information Systems Department is to develop common platforms for the whole organisation, keeping in mind as a main goal the improvement of the company's overall productivity. The intranet concept can play an important role here as it will allow to build a powerful, flexible and portable platform for all the employees and the work functions they perform: for instance not only document search and filing but also placing orders, scheduling meetings or even keeping track of working hours.

The core of an intranet is the Web system but with additional control features to limit its use internally. The Web, though, is not the only element in an intranet: there are quite a few Internet applications based on the same technology but with peculiar aspects matched to the corporate environment. Quite often, rather than proprietary network protocols (IPX, AppleTalk, SNA, etc.) corporate LANs use the Internet IP protocol, either as the single one or co-existing with some of the others. In fact, the choice of the IP protocol in the Internet has been one of the main reasons for the success of this seamless network of networks. On top of TCP/IP one finds the traditional applications such as SMTP mail, telnet remote login, FTP file transfer, News, etc., initially Unix-based but also running on multiple platforms (DOS, MS Windows, Apple Macintosh . . .). In addition to them there are now other applications for accessing distributed multimedia information (usually based on the client/server model), the star of them being the WWW based on open standards such as the hyper-text transfer protocol (HTTP) and the hyper-text mark-up language (HTML).

In the intranet we encounter the same protocols and applications restricted to local environments. Since the available bandwidth is much greater now, a faster time delay than those usually experienced (suffered) in the global Internet is perceived and, as a consequence, the overall productivity is increased. Instead of the unbearable waiting times to fill up the screen with image-rich Web pages from remote servers, in the intranet this takes place almost instantaneously. With the same Web interface, in addition to accessing to diverse multimedia information, users can transfer files, login into databases, send e-mail messages, fill out electronic forms or participate in group discussions.

The usual case is a corporation with a head office and several local branches located in different geographical zones. The alternatives to link the various LANs are: dedicated leased lines, public data networks and the Internet. The first option is straightforward and the most secure but may be too expensive to be afforded by medium- or small-size organisations. These would rather use public data networks; Frame Relay is the obvious technical option for the network layer, as the application services are based on the Internet's TCP/IP protocol suite but without any connection to that open network (in this case the IP addresses need not be official Internet addresses).

The communication option in a corporate intranet is precisely the Internet (Figure 1). LANs are then linked via Internet Service Providers (ISP); in fact there could be more than one, even in several countries or regions. The

idea then is building private IP 'tunnels' over the global IP infrastructure of the Internet, adding the necessary security elements such as *firewalls* to isolate the internal Web server from the World Wide Web and preventing unauthorised accesses from the external world. In this way any intranet user in any corporate office throughout the world can access any corporate Web server as long as he or she has the necessary access privileges.

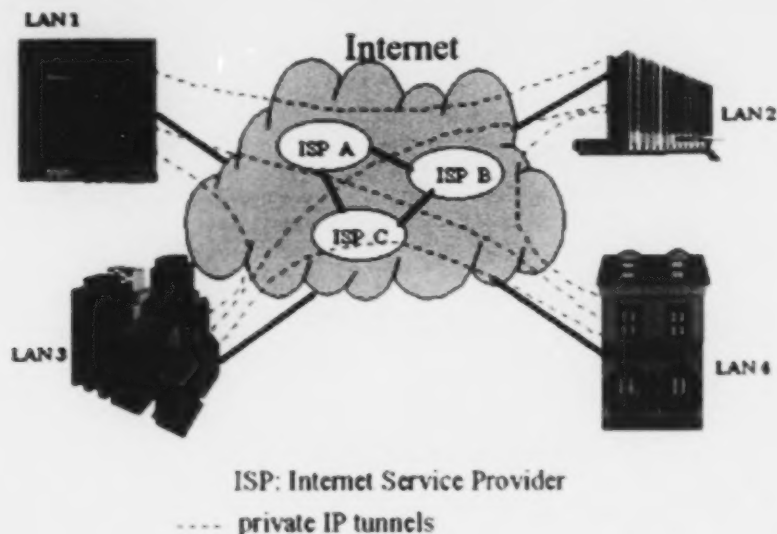


Figure 1: ISP: Internet Service Provider — private INPUT tunnels.

The last (but not least) distinctive feature of the intranet is management. That intranets use the same Internet protocols and standards does not mean that they have to follow the same (supposedly) 'chaotic' rules. In addition to security this aspect is of utmost importance for organisations that aim at building a network system to support the overall productivity, bearing in mind the annual balance sheet. According to that, access privileges of each individual and group should be carefully considered in relation to the specific jobs. Free navigation through WWW servers that hold leisure or little relevant information for the company's activities should probably be avoided by the intranet managers. Also, proxy servers and cache space should be set up in order to increase efficiency when accessing external servers frequently visited. An additional aid are the usage statistics programmes that will help intranet administrators to monitor internal use, as well as to fine tune certain parameters that may optimise the application's performance.

In summary, there are four main distinctive aspects of the intranets:

- (1) The internal use of the TCP/IP protocols in corporate LANs;
- (2) The Web as the common interface for all application services in internal networks with heterogeneous systems;
- (3) The communication of corporate LANs via the Internet (public ISP);
- (4) The management and access control procedures that have to be matched to the specific company's needs and activities.

4. Intranet applications

The use of Internet technology in private intranets, and in particular the WWW model, is profoundly affecting the corporate culture and changing the working habits presently in use. These changes take place in the different functional areas of a company, from sales and marketing departments to human resources and training units, all of them involved one way or another in the production and use of corporate information and many of them in group collaborative tasks as well. Some of the more relevant applications of this kind are (Ref 1):

- Corporate document publishing. Perhaps the most natural or immediate application in intranets as derived from the WWW publishing model. Examples of such documents are: internal newsletters, human resources guides, annual reports, price lists, product catalogues, job offers, etc. Other more restricted documents may be: meeting minutes, employee salaries, sales forecast, new product specifications and project overviews;
- Access to corporate directories and databases, for instance searching in telephone books, customer lists, bibliography, etc.;
- Workgroup applications such as project management, meeting scheduling, report writing, etc.;

- Training courses. An intranet allows easy and efficient access (when needed) to tutorials, training materials and the like; those can be enhanced with respect to traditional methods as they may include audio and video multimedia information;
- Presentations addressed to potential corporate customers. At any given time any employee concerned can have at hand the most up to date presentation material (documents, slides . . .) that can be shown at any company site (it could be more than one site);
- E-mail, in fact the *de facto* method for individual and group communication within a company. Also, submission of electronic forms to replace the traditional paper procedures;
- Software distribution. New WWW products such as Java are suited for efficient updates of software products which will reach users upon demand through the internal network. That way all users can have the latest versions which will need only to be updated in the central server.

An important aspect of the intranet technology is that the above mentioned applications used by the various departments and groups in a company are provided through a single, simple and easy-to-use graphical interface based on hypertext. The common use just requires 'clicking' on reference links and buttons: what have been termed *end-user comfortable interfaces*, better suited to the inexperienced user's needs than those powerful *graphical interfaces* (Windows, for instance) whose whole functionality is barely and seldom explored by 90% of the normal users. Top managers are a good example of that sample as they are usually more prone to simple devices that require little expertise and quick learning.

5. Intranet vs. groupware solutions

Traditionally collaborative tasks in companies have been accomplished with commercial groupware products from well-known software suppliers such as Lotus, Novell or Microsoft, to mention only a few. The intranet approach is another alternative to them. Will it replace the existing groupware applications? Before answering that let us discuss the main differences between the intranet and the present groupware solutions (Ref 4):

- (a) The intranet makes use of a technology based on open standards and protocols that run on diverse platforms (MS Windows, Unix, Apple Macintosh . . .). Groupware products are proprietary.
- (b) Intranet solutions are scalable; initial investments can be small and grow later on according to the users' needs, the required services and the available budget. Groupware solutions are generally sold as closed packages.
- (c) Building an intranet is cheaper than installing a groupware package. There is plenty of freeware and shareware software (developed by and available through Internet sites) for servers, clients, HTML editing, etc. To give some figures: the entry level of the Lotus Notes client is about \$70 per user whereas commercial Web browsers are in the \$20 rank. As to the server, the start-up cost alone for an intranet solution can be as little as \$1000, compared with \$10 000 for a Lotus Notes product. In terms of ongoing costs, intranet solutions do not require dedicated leased lines, as was discussed in Section 3.
- (d) Intranet documents are produced with the HTML language. There are a vast number of authoring tools available to assist users in document creation, many of them free through the WWW. With these tools document creation can now be the responsibility of content creators, thus freeing specialised information systems resources which are usually needed for the management of groupware applications.
- (e) Intranet does not require *synchronised replication* across multiple servers which are needed in some groupware products in order to keep consistent information for the whole organisation. Instead a single server is enough, as was mentioned above.
- (f) As an add-on value without additional cost, an intranet is a natural platform to access the outside world of the Internet.

Therefore commercial vendors are pondering whether to migrate from proprietary products to intranet solutions, or at least offering intranet compatibility. A double move is being witnessed now. Groupware companies are moving towards the intranet technology by including within their products HTTP, HTML editors, e-mail gateways, etc. On the other hand intranet technology is improving by including new features such as authoring tools (perhaps the weakest point of the intranet today as compared with groupware solutions) (Ref 5). The answer to the initial question about whether the intranet will replace the existing groupware applications is that it is not likely in the short term, at least for very specialised applications which may be critical for a company; some time and effort to implement it will be needed. In the medium or long term, those vendors are announcing fully-compatible intranet solutions which will be offered within their products. Therefore the posed question may not make sense any longer.

6. An intranet prototype. Fundesco's case

Fundesco is an organisation whose activities are centred about the IT world at large. In the late eighties a networking team designed and implemented the Spanish R&D Network known as *RedIRIS* which was managed

by Fundesco until the end of 1993. Since that time the networking people changed the initial network activities to other applications on the upper layers. Early in 1994 Fundesco set up one of the first WWW servers in Spain containing relevant information about Fundesco's goals, activities and publications, including an electronic bookshop. Fundesco's Documentation Centre had several BRS databases; thus a BRS/WWW gateway was designed and implemented to allow an easier and more friendly access to the databases. That was possible not only for Fundesco's employees but also for external users who could access into that information via the Internet. That system has been working rather successfully for the past two years and has proved to be a proper tool for information searching.

Based on the former system a new development has started in 1996. The aim is to move to a complete digital environment that includes, in addition to the possibilities of the present version for information searching, the actual contents of documents in graphical format as well as other content types such as images, voice and video-clips. As a matter of fact one has seen that documentation centres have established databases not only for cataloguing purposes but also to allow information searching by indexing and related means. This approach is scaleable but has certain limitations such as:

- it is text-oriented and not suitable for the increasing multimedia contents;
- contents is lacking of characterisation;
- user interfaces are too much dependant on specific commercial products, and not friendly enough for the normal user;
- graphical representation is poor compared with the capabilities of today's PCs and workstations.

On the other hand there is a trend to replace information on printed materials by digital formats.

Thanks to its multimedia capabilities and hypertextual representation, the Web model (and the corresponding technologies) is the obvious option to publish and distribute Fundesco's corporate information. Document contents can thus be represented in a standard way according to the MIME (*Multimedia Internet Mail Extensions*) conventions used in the WWW model. The goal is to develop a system in order to automate the process shown in Figure 2. Paper will be progressively suppressed and, at the same time, other information formats such as sound, images and video will be incorporated. The system is highly portable as it is based on open standards (Unix and intranet), so that it can be adapted both to small documentation centres and to large libraries.

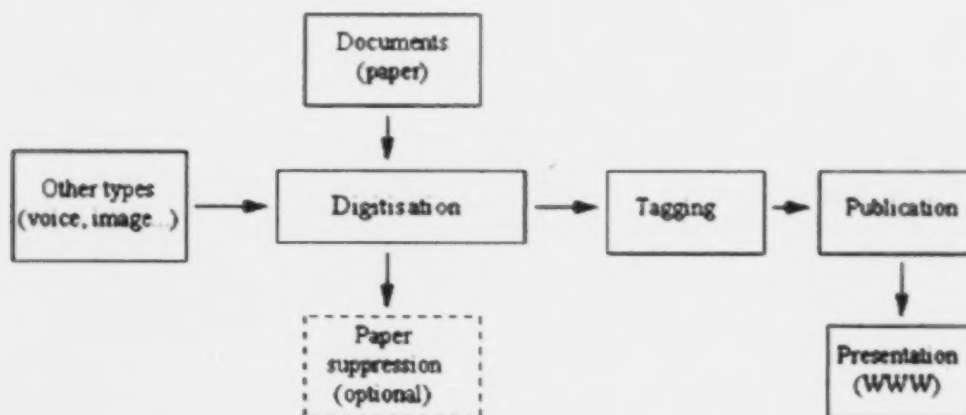


Figure 2

Figure 3 shows the different parts and the overall scheme. The *input interface* will capture the analogue information to convert it to digital format. It is composed of a high-resolution scanner and a Unix workstation connected to it. There is also a digital audio system and an analogue video card which can digitalise and MPEG compress. The tagging process, that is the association of contents (to be stored in a repository) and the corresponding documents, is accomplished at the workstation. Tags are objects that point to the document contents in the repository; they also include additional characteristics needed to present each document. The other part of the system is the *access interface* to the information. It is a BRS database system and a WWW server, both parts linked by the BRS/WWW gateway discussed above.

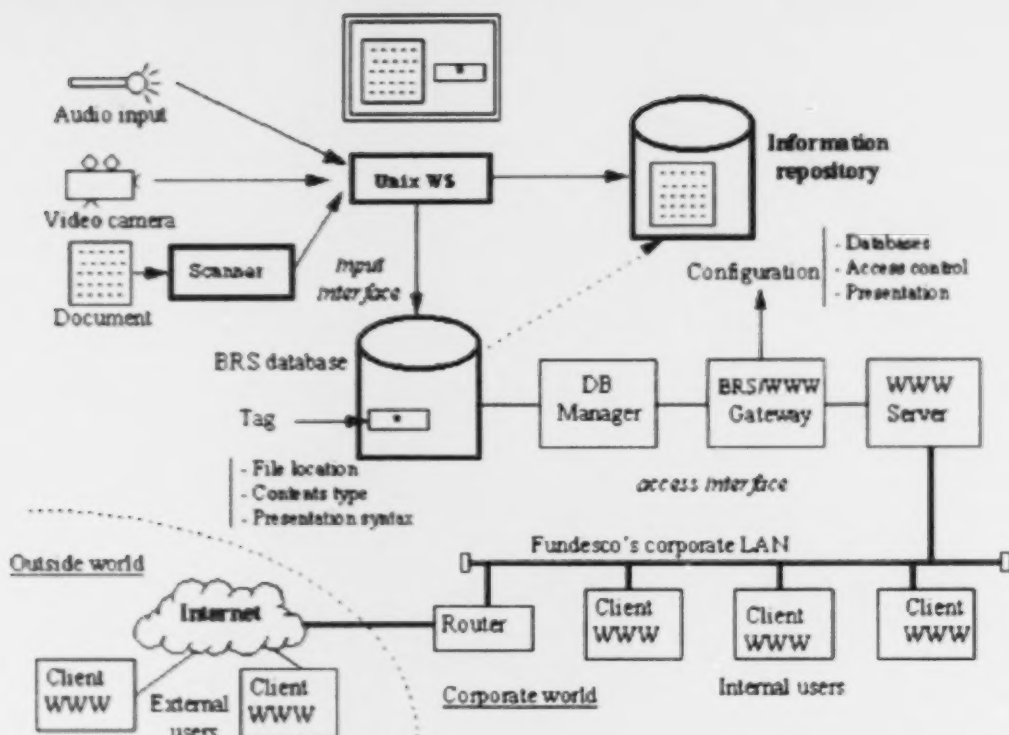


Figure 3

When a document is scanned the workstation captures it in digital format and stores it in the repository; a tag is created and inserted in the database where the bibliographic reference (as it was previously catalogued) is located. Therefore document and contents are related by the tag, although they reside physically in different parts. End-users interact with the database system via their WWW client and the gateway. Thus they need not be concerned with BRS native commands or other implementation details. At the gateway there are access control mechanisms to the database for Fundesco's employees. Documents can be searched via WWW user-friendly forms; once the required document is located users can visualise parts of it (pages), as well as other types of information such as video, audio, etc., and certainly the textual information corresponding to the original catalogue record.

In summary the system functionality is such that:

- in addition to the original catalogue record, tags show where the digitised file is located, the type of contents (text, image, voice . . .) and the presentation syntax;
- e-mail and personal Web pages of authors can be included as hyperreferences in the multimedia documents;
- home pages of document publishers can also be included in the same fashion;
- the multimedia archive system being built can be accessed through WWW hypertext, and the amount of storage space for paper can significantly be reduced;
- inexperienced users (in practice most of all the employees) have at hand a common and friendly interface;
- the system is highly portable for various platforms and systems, as the implementation is based on open standards and proven technology (Internet, intranet).

The first half of the project (March–August 1996) encompasses six months. Goals for this phase are to develop a prototype for the manual tagging of contents, the creation of tags and the manual insertion of those into the database. The digitised information will be published automatically through the WWW, as it was originally implemented but adding now the ability to handle tags (for multimedia contents) with other new features such as more flexible mechanisms for configuration, security and presentation. The second half (September 1996–February 1997) will yield a true production system. At that stage tags will be generated automatically and inserted into the database system, thus allowing instantaneous publication of documents as soon as they are scanned. It is intended too to develop a generalised interface to interact with different database systems

(documentary or other types) to be used by documentation centres, libraries and the like.

This project is an example of how an organisation can use the existing Internet technologies to build up practical intranet solutions. Fundesco's employees will have simple and easy access to the multimedia information of the Documentation Centre from each desktop on the corporate TCP/IP LAN through a common WWW browser.

José Barberá
Fundesco
Networking Department
Plaza Independencia 6
Madrid 28001
Spain

References

- [1] JSB Computer Systems Ltd, *The Intranet. A Corporate Revolution*,
<http://www.intranet.co.uk/intranet/intranet.html>.
- [2] Process Software Corporation, *Intranet — A Guide to 'Intraprise-Wide' Computing*,
<http://www.process.com/news/intrawp.htm>.
- [3] Barberá, J. and C. Blázquez (1996) En los dominios privados de Internet, *Comunicaciones World*, No. 101, May.
- [4] JSB Computer Systems Ltd, *The Intranet Vs. Groupware Solutions*,
<http://www.intranet.co.uk/intranet/sidebars/groupware.html>.
- [5] Martínez, I. and I. López (1996) Centrados en 'la red', *Comunicaciones World*, No. 101, May.



Electronic current awareness in the corporate environment

Elizabeth Sale

KPMG, UK

Abstract: *Timely and authoritative news on companies and industries is vital in today's competitive business environment. Information departments have traditionally played a role in providing current awareness services (CAS) to deliver the information required. Technological developments such as groupware and Web technology have opened up new opportunities for the delivery of information direct to the end-user's desktop. These advances have meant that suppliers are now producing a new breed of current awareness services, termed alerting services, which automatically filter newswires and download 'profiled' stories to the user's desktop each day. Automated CAS enables a more efficient and effective way of disseminating external information throughout an organisation, leading to a more informed decision making culture. The vast array of services available and the costs of subscribing and operating a system demand a thorough analysis of the system's features and the needs of the user. A pilot study within KPMG identified three key criteria that must be considered: compatibility with the existing IT infrastructure, the delivery platform and the content of the information being filtered.*

Keywords: Current awareness services, automated alerting services, end-users, Lotus Notes, intranet

1. Introduction

This paper reviews the changes that are occurring in current awareness provision through a consideration of the preliminary results and research findings from a current awareness pilot taking place in KPMG. The results collected from this study represented the empirical element of a Master's Dissertation and the research findings have led to a further pilot being established so that the most appropriate method of CAS provision for KPMG's employees can be determined. The paper focuses on the service levels demanded by end-users in order to perform daily tasks effectively and in the final analysis to achieve the goals of the organisation.

The key feature of each of the services reviewed is that information is delivered directly to the end-user. One of the central thrusts of the study is the shift of information retrieval from the intermediary to the end-user. A review of the literature on current awareness demonstrated how this shift has occurred and how alerting systems have evolved. Three external influencing factors have been isolated:

- the advances in information technology and networks that have affected information retrieval and provision;
- the recognition of the importance of information and the changing working patterns of organisations;
- the rise of information products aimed at the end-user.

2. Development of current awareness systems

2.1. Principles of current awareness services (CAS)

Our study began with a review of the main principles of current awareness services. Library and information services in industry developed CAS to help keep individuals aware of literature that affects their subject area as it is published. The quantity of information available means that an effective current awareness service is invaluable to those who depend on inward information.

By establishing a profile for the end-user, the Library and Information Service (LIS) develops an understanding of its client's needs and establishes a relationship in which those needs are more deeply explored and continuously refined. The development of such a service also represents a marketing opportunity for the LIS to demonstrate the full range of value adding services it provides.

Whitehall (Ref 1), Rowley (Ref 2) and Blick (Ref 3) contend that the key features of a successful service are:

- (1) detailed knowledge of the client's interest;
- (2) appropriate coverage of sources;

Online Information 96 Proceedings

Page 195

BEST COPY AVAILABLE

- (3) timeliness of delivery;
- (4) good selectivity techniques so that relevant material is retrieved;
- (5) ease of use and access to the information;
- (6) cost;
- (7) customer interaction and feedback;
- (8) presentation of output in a readable manner.

Additional research by Rowley (Ref 4) concluded that the critical success factors for an effective service are that it should:

- (1) keep the information user up to date;
- (2) save project time;
- (3) aid in the creation of new ideas;
- (4) save time and money spent on journals;
- (5) reduce the paper flow in the organisation.

2.2. The impact of information technology on current awareness services

The impact of information technology (IT) on the development of current awareness services has been extensive and continues to be significant. The development of the information industry has moved hand in hand with the development of information technology. Communications and networking technologies combined with powerful, easy to use personal computers have given the end-user access to vast banks of information. The evolution of electronic information services, and in particular the development of the Internet, has made available a wide range of information sources. Electronic information is not better or superior to printed information, it merely increases and extends access to information sources. However, the ease of redistribution of electronic information introduces the risk of breach of copyright. It is therefore vital that an understanding of the limitations on the legal distribution of material is understood by those who handle it. Developments in electronic publication have led to the electronic storage of information and its rapid retrieval. Data communications is the fastest growing field of telecommunications, allowing considerable amounts of information to be delivered from a remote host to the desktop. Many organisations are now having seriously to consider adapting their information services, and the way in which they operate, to take advantage of the opportunities presented by the Internet. With IT breakthroughs, it should be remembered, 'Human intervention is needed to establish just what an item is about; the job which involved a semantic judgement of a document's "aboutness" could not be left to the computer entirely' (Ref 5). However, online searches 'are not only more productive but also more effective than manual methods' in scanning vast amounts of sources and distributing information quickly (Ref 6).

2.3. Learning organisations

Information is now recognised by organisations as a valuable asset (Ref 7). Organisations are embracing new technology and their employees are using IT applications more frequently and for more tasks. As Paul Waddington (Ref 8) noted, IT is changing our lives and working practices. Advances in telecommunications have changed the way in which businesses operate. The increase in network infrastructures has meant that the virtual organisation is now clearly emerging. Working from home is a reality in many organisations. This has led to the need for a 'virtual library' that can deliver information when, where and in whatever format is required. Today's LIS must be able to deliver services direct to the end-user's desktop. The shift from manual to electronic technology has led to a new breed of information centre. The market for information products, once restricted to the information intermediary, is now finally reaching the end-user. Information providers have always regarded the end-user market as the Holy Grail but are only now breaking into it. The market place is at last being driven by the needs of the end-user. The challenge for producers now is to meet the needs of the end-user because 'for every professional there are 100-1000 non experts' (Ref 9). The barrier to this market in the past has been technology. Now with all of the advances in IT these barriers have been removed — the majority of users will not think twice about obtaining information electronically in the future. Alerting services can be viewed as the next wave in the sea of changing information.

3. KPMG and current awareness

3.1. KPMG

Where does KPMG stand in this picture of new wave information products? KPMG is an international firm of accountants and management consultants serving local and global clients. There is a great need to identify and exploit new business opportunities as they emerge and to anticipate the needs of clients. Partners and staff must be kept abreast of the regulatory developments and business trends that are likely to impact on their clients. For this reason it is essential that everyone in the organisation is kept informed about the financial and business environment, and have access to information sources which are accurate and timely. The Library and

Information Service supports KPMG's 38 UK offices. One of its key responsibilities is choosing information products that equip the organisation with the knowledge to be flexible, creative and innovative. The LIS's strategy focuses on developing services which demonstrate product relevance and cost-effectiveness. The Library Manager needs to anticipate and review the changing information requirements of the organisation and to provide a service that matches these needs. One of these needs is the development of automated alerting services.

3.2. Background to the CAS pilot

The LIS has provided a comprehensive range of services over the last ten years, including circulation of contents pages, manual clipping services and distributing journals. The introduction of electronic information products, particularly Reuters' Textline, meant that a computerised electronic cuttings service could be implemented for regular searches (i.e. Selective Dissemination of Information — SDI). The number of requests for this service has increased each year to such a degree that it took an information specialist two days each week to process the SDIs and distribute them throughout the organisation. There are advantages and disadvantages to an SDI service. The LIS has complete control over the running of the SDI search. Searches can be established, changed or deleted at the information specialist's command. Conversely, maintaining the service requires the continuous processing of information by the information specialist. The demand on LIS's time meant that to extend the service to run on a daily basis was not a viable option without recruiting a full-time dedicated researcher to supply SDI information. For this reason the service could not provide very current information, nor could it cope with the volume of people who wanted to benefit from the service. In short, the system was at breaking point. Automated electronic alerting services allowed the LIS to provide the user with daily information without impinging on the information specialist's time.

3.3. The way ahead for the Current Awareness Pilot

In March 1994 an evaluation of all the existing CAS services in KPMG took place. A questionnaire was distributed to the CAS users to establish their views on the service. The results of the survey highlighted a number of strengths and weaknesses:

Strengths

- tailored to a specific audience;
- information produced being highly relevant to the end-user.

Weaknesses

- insufficient to meet actual demand;
- labour-intensive;
- costly;
- duplication of services occurred through lack of co-ordination between departments.

We concluded that a more coordinated and effective way of alerting people should be established. Three options for future current awareness provision at KPMG were identified: centralise the existing services, introduce a new comprehensive service, or replace manual gathering systems with an automated system.

We decided to carry out a focused review to examine whether the introduction of a fully automated system would reduce the labour-intensive nature of the existing services, permitting redeployment of personnel and the maximisation of resources. We sought to test the hypothesis that an automated service would cope more readily with the growing need for information throughout KPMG.

Our review consisted of three stages:

- (1) review questionnaire results to determine how end-users felt about CAS (developing a blueprint for the type of service required);
- (2) the selection of individuals and services for the implementation of the pilot;
- (3) the distribution of a questionnaire to obtain feedback on the pilot.

We will now consider each of these in turn.

3.3.1. Review questionnaire

The results of the questionnaire established end-users' requirements for CAS provision and allowed us to formulate a strategy to ensure provision of a service that would meet KPMG's current and future information needs. The operating criteria for the system were:

- the service must provide very specific information, directly matching the user's profile;
- the information must be accurate, timely and relevant;
- a maximum of 20–25 headlines should be supplied initially, with the user being able to access the full text of each article immediately;
- the recipient must be able to print and view the information immediately and locally;
- the information provided must be in the public domain and must be news data;

- the systems on offer must take into account copyright restrictions but should allow information to be stored and distributed as necessary;
- text retrieval of archived data should be possible;
- the service must be flexible with the ability to change, add or delete search profiles on a daily basis.

We then assessed the market for electronic CAS, evaluating the systems on the following criteria:

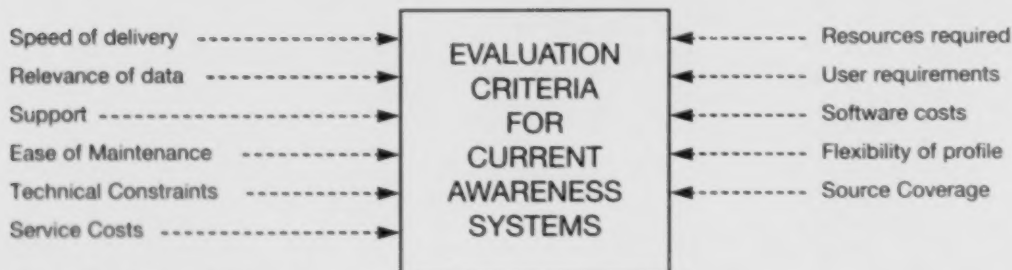


Figure 1: Evaluation criteria.

With these criteria in mind a number of service offerings were assessed to find the most suitable service for the pilot study. Our objective was to find the most relevant and cost effective CAS for KPMG.

3.3.2. Implementation of the CAS pilot

The pilot was divided into two phases. Phase one involved implementing, testing and reviewing the trialed CAS:

- identifying the potential benefits of the service;
- identifying a representative pilot group of users and their requirements for the service;
- implementing a trial of two commercial services (not simultaneously) considered likely to meet the requirements of the users;
- seeking feedback from the users;
- documenting and reviewing the results.

Phase two at the end of the six month trial period consisted of:

- collecting feedback from the pilot users via questionnaires;
- assessing the potential costs and benefits of the services;
- qualifying the technical approach, scalability of the service and its impact on infrastructure;
- documenting the results of the pilot and recommending future actions.

3.3.3. Questionnaire and feedback on the pilot

Although the pilot users had continually fed back their views on the CAS via meetings and questionnaires, at the end of the pilot a comprehensive questionnaire was sent out to users to capture their opinions on automated systems. The questionnaire covered issues such as, what preference does the user have with regard to various service details? what prompts the user to make more frequent use of the service? is the information content paramount? and, does the user need daily information feeds? The information industry seems to be tailoring products for real-time feeds but is the end-user ready for such a change in working habits?

The questionnaires were analysed and the key issues were isolated:

- (1) the source coverage of the system was deemed to be the most important criterion for choice, followed by speed of delivery;
- (2) flexible output options were essential as some favoured printing, others viewing on screen;
- (3) users need to be educated to take responsibility for their information needs and to understand that they will have to take time out of their working day to give feedback to the product suppliers. If there was not much interaction with the suppliers for building and maintaining the profiles, the service proved inadequate;
- (4) there was considerable enthusiasm for automated CAS from the end-users but the means of delivery and scope of support for the service were considered to be of great importance. One suggestion was the provision of a helpdesk for any queries that arose;
- (5) one hundred percent of the respondents wished to continue receiving an electronic alerting service, although more than half of this number said that commitment to a subscription would be dependent upon cost;
- (6) service providers need to gain the user's trust and confidence before their products can be regarded as an essential business tool. One of the respondents commented that they used the service weekly but

would use it daily if they had more confidence in the reliability of the sources, and if they felt that the system was picking up everything of relevance.

4. The delivery platform

Our initial evaluation criteria for CAS services concentrated on scope, content, currency and ease of use. However the results of the trial demonstrated the importance of the delivery mechanism. We needed to consider the following:

- computer network infrastructure;
- compatibility of the software/hardware;
- user training;
- cost.

One of the major issues of automated electronic current awareness services is the trend of shifting responsibility for information retrieval from the library directly to the end-user. This means that the delivery platform is crucial. Which is better, 'push' or 'pull' services? Should the system run in the background and alert the end-user when a message arrives or should the users enter the system at their convenience? If a service is to be implemented successfully, the product front end needs to be intuitive and uninhibiting so that the user is not faced with a barrier to retrieving information. The mechanics of information retrieval must play a minimal role in the user's daily routine, as the end-user is only interested in the 'message' not the 'medium'. A successful delivery platform depends on the delicate balance of many variables. The ideal for CAS provision is that information should be delivered to the right person at the right time, in the right format and for the right price (Ref 10): finding a delivery platform that can offer such a service is no easy task. Information suppliers are constantly trying to improve their delivery mechanisms: they realise that failure to keep abreast of technological developments may result in the downfall of their service. Current Awareness providers are joining forces with information disseminators so the two can provide an efficient service. Today the results of an information search can be delivered by post, fax, modem, e-mail, groupware or the Internet/intranet, or any combination of these. Suppliers need to be able to offer all these delivery mechanisms in styles that can be incorporated into different organisational infrastructures. What may be good for one company may be completely inappropriate for another. For the initial KPMG pilot study Lotus Notes groupware was chosen. At this stage intranets were very much in their infancy and the Internet was seen as posing possible security hazards. Lotus Notes was chosen on the basis that it allowed delivery of articles, e-mail and the construction of databases, and did not pose the security hazards inherent in the Internet. The LIS worked very closely with the IT department. An additional feature of the trial was that it helped establish the value and potential benefits of Lotus Notes to the LIS — an extra benefit from the pilot.

Lotus Notes was considered particularly suitable for CAS delivery. Filtered information is delivered from the external supplier to the organisation's server by replication of databases so that the supplier is able constantly to update the organisation's server. The end-user can then access information held on the Notes server through the Notes Workstation (i.e. the individual's PC). Information management is one of Notes' core strengths. One reason for its success is the structured manner in which information is stored, in documents, forms and views. *Documents* store the information content of a database which can include text, graphics, tables and sound. *Forms* are templates that help to create documents. *Views* are lists of documents stored or categorised in a variety of ways to make it easy to find information. Views are comparable with table of content for the database. The ability to search the CAS database using views allows the user to retrieve information in a variety of ways, for example by date, publisher, source and so on. Smart icons facilitate the manipulation of the screen. An additional feature of the Notes database is that it allows the searching of views by Boolean terms.

Organisations need to determine which infrastructure will best suit their working patterns. After the initial pilot we reviewed alternative delivery platforms to Lotus Notes, in particular the Internet and intranet. The development of World Wide Web technology has added another dimension to information provision, offering a multimedia environment at very low cost on multiple platforms. Intranets for CAS represent a cost effective delivery option which should be considered particularly if Lotus Notes is not already an established application within the organisation. The use of internal Web sites removes the security risks of the Internet. It is apparent that an open system is a more suitable system for CAS delivery than a closed system. KPMG is currently piloting CAS delivery over the intranet.

5. Results

The results of our pilot study demonstrated that three organisational factors need to be given serious consideration before an alerting service is established:

- IT infrastructure harmonisation;
- the likelihood of users incorporating information retrieval into their working practices;
- the delivery platform.

All of these factors are equally important to the acceptance of the service by the end-user. In addition to these

main criteria, the following issues were identified throughout the pilot study:

- the delivery platform is a major factor in the effective utilisation of the CAS systems. If the system is not intuitive to use, requires a lot of training and support or is very expensive to implement then the CAS will not give the desired results to the organisation. For the first pilot within KPMG Lotus Notes was used as the delivery platform: the second pilot is using the Intranet;
- good contact and communication both externally with the suppliers and internally between departments, particularly the information technology department is essential if a service is to be introduced successfully;
- clear coordination of all information systems is vital if information is to flow effectively throughout the organisation;
- the end-user is key to the successful future of electronic alerting services. If services are to be used to their full potential then the user needs to become more proactive in his or her request to receive relevant information;
- users need to be very clear about what kind of service they require. In busy environments information needs on specific topics change very rapidly (often day-by-day). Users need to be confident that the system is appropriate for their needs;
- the strengths of sources of each system need to be carefully understood before selection;
- the information specialists should coordinate the service to the users, alter profiles, train the users and offer support when needed. When users were asked whether they preferred to use the supplier or the library as primary contact for an alerting service, an overwhelming majority of users replied that they preferred to use the library because they felt that the library had a greater understanding of their information needs;
- the addition of automated alerting services will allow the library to provide information to a greater number of users more frequently, efficiently and effectively.

6. Conclusion

Whilst end-user access to information is still in its infancy, it is rapidly becoming a meaningful and viable option. Alerting services are just one facet of the changes in information provision. Changes and developments in information products will impinge on the way the librarian/information professional operates. As more people become aware of the importance of information, we will see the metamorphosis of information professionals' roles. 'A librarians job used to be building and maintaining a collection of books, now the service has shifted to service provision orientation — user needs orientation' (Ref 11). Royan argues that in the midst of these changes, information professionals have an opportunity to develop new roles:

- (1) Evangelist — persuading people to use services;
- (2) Gatekeeper — registering the user;
- (3) Facilitator — bringing the resources together for the end-user;
- (4) Cartographer — documenting the services available and the paths;
- (5) Trouble shooter — helping the end-users when they get stuck;
- (6) Trainer — showing the end-user how to use the products;
- (7) Packager — sorting the interfaces (Ref 12).

LIS staff are faced with exciting opportunities. The end-user need not be a threat. Information professionals have expert knowledge of information systems and valuable, unrivalled experience. There is no reason why the end-user should replace the intermediary — the two can be complementary. The LIS should take the lead in identifying and introducing appropriate new information products. Doing so will ensure that an effective information strategy is developed and implemented so that the needs of the new knowledge workers are met in full.

Elizabeth Sale
KPMG
1-2 Dorset Rise
Blackfriars
London EC4Y 8AE
UK

References

- [1] Whitehall, T. (1986) *Practical Current Awareness Services from Libraries*, Gower Publishing Company Ltd, Chapter 1, p. 1.
- [2] Rowley, J. (1992) Current awareness and the online hosts, *Proceedings of Online Information '92*, p. 251.

- [3] Blick, A. (1986) The choice is yours, in *Practical Current Awareness Services from Libraries*, Gower Publishing Company Ltd.
- [4] Rowley, J. (1992) Current awareness or competitive intelligence: a review of the options, *Aslib Proceedings*, **44**(11/12), 367.
- [5] Pfaffenberger, B. (1990) *Democratising Information: On-line Database and the Rise of End-User Searching*, GK Hall and Co., p. 38.
- [6] Naber, G. On-line versus manual literature retrieval, *Database*, **8**.
- [7] The Hawley Committee (1995) *IMPACT Checklist and Explanatory Notes: Information as an Asset*, KPMG.
- [8] Waddington, P. (1995) Information as an asset: the invisible gold mine, *Business Information Review*, **12**(1), July, 26.
- [9] Rennie, J. and E. Hamilton (1993) Electronic business information towards 2000: on-line providers look at the future, *Aslib Proceedings*, **10**(2), 58.
- [10] Cropley, J. (1989) As you sow, so you shall reap: understanding the value of information, *Aslib Proceedings*, **41**(11/12), 326.
- [11] Royan, B. (1990) A practitioner's view of self-service information systems, in P.T. Bysouth (Ed.) *End User Searching — the Effective Gateway to Published Information*, p. 79.
- [12] Royan, B. (1990) A practitioner's view of self service information systems, in P.T. Bysouth (Ed.) *End User Searching — the Effective Gateway to Published Information*, p. 81.



Intelligent agents: information strategies for the Information Society

A.J. Garcia-Sierra

BT Electronic Commerce Innovation Centre, Cardiff University of Wales, UK

Abstract: *The ever-increasing requirement for enterprises to communicate more quickly and more efficiently is evident in a time of decreasing trading barriers in national and international arenas. There is an increasing requirement to acquire, exchange and disseminate the increasing mass of information continually reaching our desktops. An information strategy will be the single most important asset of an innovative enterprise within the next 10 years. As identified in the Bangemann Report in 1994, the Information Society will be a revolution which 'adds huge new capacities to human intelligence and ... changes the way we work together ...' (Ref 1).*

The workplace of today is increasingly being overloaded with information. The concept of intelligent information agents has been widely prescribed as a solution to this problematic overload for some time — agents have effectively been labelled as any localised software which does not reside in a mainframe (Ref 2). However, negotiation is a core part of the software functionality and agents must exhibit some form of intelligence (i.e. they can say no to a request and re-route their queries).

Careful design and development of such agents could well form the basis for tackling the Information Revolution.

Keywords: intelligent agents, Internet, information society, flexible information working

1. Introduction

This paper briefly looks at the UK Government's Information Society Initiative which has been fuelled by the realisation that information is the key component of the on-going information revolution. Information superhighways are exemplified as the high-level path to which companies should be aspiring. In order to follow such a path there is a requirement for an enterprise-wide information strategy. The paper then concentrates on the role of intelligent information agents in providing future generations with a superior course for flexible information working.

2. The Information Society Initiative

A need for a nationwide drive to improve the competence of and access to sources of advice and products to help businesses harness information and communications technologies (ICT) has become clear in recent years. The need for suppliers to orient themselves more towards the needs of businesses, particularly small to medium-sized enterprises (SMEs), has also been growing in evidence. In these respects the UK Government launched the Information Society Initiative (ISI) in February this year as a partnership between industry and Government to help British business thrive in the emerging information-based economy. The Initiative is a four-year programme, aimed to conclude in the year 2000 (Ref 3).

At an almost exponential rate we are seeing computing power roughly doubling every year. This means a brand new PC on your desktop is twice as capable as one bought just twelve months ago — this incredible processing power is being made available at cheaper and cheaper prices in real terms. Perhaps the most important facet to note is that this increasing power, not only of computers but also of communication networks, is used primarily to get hold of information we want.

The industries that create and market these products and services, from telecommunications and broadcasting to information technology and publishing, are increasingly converging to form a single electronic information society. The intelligent use of information, or knowledge, is the key to economic and social development.

The ISI has responded to the competitive challenge rapidly emerging — a challenge in which all components of industry need to respond; to adapt to new ways of thinking and working; to invent and innovate with new products and services; to adapt to new structural relationships with suppliers, customers and other trading partners; and to source new customers in flourishing markets.

3. Information superhighways

In December 1993 the European Council requested that a report be prepared for its meeting in June 1994 by a group of prominent persons on the specific measures to be taken into consideration by the Community and the Member States for the infrastructures in the sphere of information. This report subsequently became known, in the light of his Chairmanship of the High-Level Group on the Information Society, as the Bangemann Report (Ref 1).

Although a very generalised document, specific aspects of the report have laid the foundations for the mass utilisation of information in all its various forms for the near- to mid-term future — essentially what we now know as the information superhighways.

4. Information strategy

An enterprise-wide information strategy could quite simply become the single most important facet of a company operating in the next millennium. Whereas we are all fully conversant with the concept and reality of strategies for production and manufacturing, purchasing, sales and marketing, finance, computing and more recently information technology or systems, and so on, are we not quite so clear that universal to all of these is the core requirement of collecting, collating, disseminating and exchanging information? So, why do companies not seem to be doing just this?

The bare requirements for an enterprise-wide information strategy are unambiguous. Information needs to be got to the right places at the right times. Further, decreasing barriers for international trade are becoming evident around the globe on a daily basis. This must be realised by enterprises who truly wish to incorporate innovation into their company — use the information that is available out there!

But what exactly constitutes an information strategy? For decades companies have been organised according to separate, and often disparate, functions. However, organisations must begin to realise that paper-based and machine operating environments are no longer the dominant domain. The influx of desk-top computers in the last 10 years has dramatically changed the way organisations operate. It is clear that the single most important asset of any organisation is information (perhaps it is time to change the old adage 'it's not what you know, it's who' to something like 'it's not what you know, it's who you know, and who they know, and so on'). The key point being that communicating information is vital in this day). Therefore it is unequivocally logical that an information strategy should be at the top of an organisation's agenda, and not Sales & Marketing, nor Distribution, nor Purchasing, nor Quality, nor Manufacturing, *ad infinitum*.

Each of these traditional functions will become part of the information strategy. Such a strategy should allow organisations to focus on what are the most important communications flows internally and externally; who are the key contacts; which are the key channels; what are the best sources; where are the biggest bottle-necks to success; what makes the internal (and external) supply chain flow and so on.

5. Role of the Internet

As a desk-top accessible medium of information transference, the Internet not only offers an extraordinary wealth of information for those searching (sourcing/input) but also for dissemination (output) and additionally as an exchange medium. This continually increasing medium is the single most important factor in developing true information superhighways. In fact according to John Patrick, Vice President of Internet Technology at IBM, the Internet 'is the information superhighway' (Ref 6). The Internet offers an infrastructure on which an information strategy may be adopted and developed. However, a pre-requisite and possibly the killer application of the future Internet is the concept of intelligent information agents — software which is utilised to deliver such strategies efficiently.

As the Internet, and in particular the hyper-linked graphical element known as the World Wide Web (WWW) has grown, so the number of documents on it has naturally increased (Ref 7). Mauldin estimates that the WWW consists of 32 billion bytes (29.9 gigabits) of document data (Ref 8). Even allowing for inaccuracies in the sampling approach and a substantial element of duplication (in 'mirror' sites), there is still a vast amount of data available online. The more successful the WWW, the greater the problem of information and resource discovery.

It is on these thoughts that the remainder of this paper is based: intelligent agents accessed via a simple desk-top interface, just as the Internet is accessed via easy-to-use graphical browsers.

6. Intelligent information agents — overview

'The computer language for agents will be completely safe and very flexible. It may even read your mind occasionally' (Ref 9).

The concept of intelligent information agents has been widely prescribed as a solution to the problematic infor-

mation overload for some time — agents have effectively been labelled as any localised software which does not reside in a mainframe (Ref 10).

Techniques in artificial intelligence (AI) in the past have tended to 'use knowledge about the interests and priorities of people to perform routine organisational tasks such as automatically screening, directing, revising and responding to information' (Ref 11). There are essentially two types of approach behind AI which drive agents:

- 'symbolic AI' approach or 'deliberate thinking' paradigm (Ref 11);
- 'nouvelle AI' approach (or behaviour-based AI) or situated agents (Refs 11, 12).

The former approach patterns itself on a mental mode of intelligence — that is symbolic representations of knowledge, usually configured as intelligence residing in the machines as a set of structures. The agent is essentially meant to be governed by logic. This approach is clearly computing resourceful and requires a great deal of prior knowledge.

The latter approach make less assumptions of internal attributes and does not build elaborate knowledge patterns. These agents become engaged dynamically with other agents and as such build their own 'picture' of events — this picture then assists in activating and completing tasks. This implies by its very nature that learning is a prerequisite for these agents. The learning process is achieved through dialogue with other agents, host computers and subsequent recording.

There appears to be growing acceptance, within and outside of intelligent agent theories, of the idea of behaviourally-based agents which literally sit in an environment and learn whilst there, rather than agents which are pre-defined with intelligent structures but are less able to learn for themselves (Ref 10). In October this year the Intelligent Systems Group of British Telecom's R&D Laboratories, in Martlesham Heath, published a set of papers in the *BT Technology Journal* (Ref 13) in which this concept of behaviourally-based agents was fully endorsed (to be referenced).

It is the inherent levels of intelligence previously discussed that must be incorporated into the careful design and development of such agents. It is the opinion of the author that these agents will form the basis for tackling the Information Revolution — in other words the 'killer application' of the Internet. But what must be remembered is that agents need effectively to be off-the-shelf products which will assist the management of the ever-growing 'software and information mountains' not only for the cash-rich multinationals of this world but also the thousands of SMEs.

Considering the case of companies which cover large geographical areas, and those which last for a number of years, there is clearly the need to develop agents which can cope with collaborative activities so that inappropriate decisions are avoided. In building agents within an infrastructure which actively supports storage, retrieval and delivery of pertinent information to appropriate decision-makers, activities can continue successfully across processes despite turnovers in personnel (Ref 11).

7. The requirements of intelligent information agents

But what exactly are the requirements of intelligent information agents? Generally, agents must be able to:

- identify which other agents to communicate with;
- establish a reliable communication channel with other agents (and back-up if necessary);
- identify and be able to use a protocol for the ensuing dialogue;
- identify and use a common language to exchange information knowledge;
- know what terms within the language to use to guarantee that the other agent(s) will interpret the expressions in the same way (compare this with 'classical electronic data interchange');
- know how to handle inconsistent information and the eventual mismatches that arise from different languages, views, translations and so on around all of the thousands of distributed computers in the world which are hooked up to the Internet.

These requirements begin to proliferate in magnitude of potential issues, and benefits, as an increasing number of agents are involved in information transactions.

The Media Laboratory at the Massachusetts Institute of Technology (MIT) is conducting a number of projects and experiments to develop the evolution of intelligent agents further (Ref 14):

- *Software Agents* project is applying AI techniques to the field of human-computer interaction;
- *Agents that Reduce Information Overload* project is attempting to deal with the problem of information overload. Software agents are being developed that make personalised suggestions to a user for items the user may want to select (news articles, videos, music, television shows, etc.). This project employs two different techniques: content-based filtering (detection of patterns among the items liked or disliked) and collaborative filtering (detection of patterns among different users and making recommendations to people, based on others who have shown similar tastes);
- *Yenta — Matchmaking Agents* project is developing a software agent that finds people who have never met, but share similar interests, and introduces them to each other. Such introductions can form interest groups and coalitions automatically and can be used to locate someone knowledgeable in a particular area;

- *Remembrance Agents* project is addressing the common human fault of bad memories (unlike humans, computers almost never forget the information they store!). This project is creating agents that help augment human memory by logging everything a user does and all the information which passes through the user's hands, thus helping a user remember some information based on the content or the context of the situation;
- *Kasbah: An Agent Marketplace for Buying and Selling Goods* project is helping to realise a fundamental transformation in the way people transact goods — from requiring constant monitoring and effort to a system where a software agent does most of the work on the user's behalf.

Each of these projects has the potential to reform everyday working practices fundamentally by being applied to business applications.

8. Commercial intelligent agent scenario

The following scenario looks at the three main basic 'processes' of a commercial enterprise, namely purchasing, producing and selling.

8.1. Purchasing

As a major function of any organisation, purchasing should ideally be efficient and effective. One service freely available on the Internet today can assist in this area — *URL-minder* sends users an electronic mail message when a specified Internet site changes (i.e. is updated or edited in some way). Although the service is currently somewhat sporadic, and is best suited for tracking special-interest Web sites that change infrequently, it is a clear indication of how buying could be fundamentally altered in the future.

For example, a similar agent could operate freely on the Internet tracking Web sites of known (and newly formed) suppliers of goods and services which a company may wish to purchase. If a supplier decides to offer discounts or some kind of 'deal' then the purchasing agent could alert its user immediately. Intelligent agents could also be sent out on open networks to gather market intelligence, vet potential suppliers, barter on behalf of a company and even negotiate unique contracts.

The aforementioned MIT Kasbah project provides an excellent overview of how a user wanting to buy or sell a good can create an agent, give it some strategic direction and send it off to the agent market place. The Kasbah agents proactively seek out potential buyers or sellers and negotiate with them on their creator's behalf. The agent's goal is to make the 'best deal' possible, subject to a set of user-specified constraints, such as a desired price, a highest (or lowest) acceptable price and a date to sell (or buy) by. These agents need to be intelligent enough to perform well in a complex, dynamic marketplace and a key feature of Kasbah is that it is open to adding new types of agents which use different selling strategies.

8.2. Producing (processing)

Once the pre-contract stage has been passed and contracts formally set up, all with great assistance from agents, there is the opportunity for Intra-agents (that is agents operating exclusively within a company) to monitor production across the various business functions. For example, a manufacturer might have Intra-agents residing within internal networks which follow the flow of information pertaining to a product's development and life-cycle. Information which needs to be passed through the company could be initiated automatically so that effectively users receive up-to-date information that is only relevant to them.

The Yenta — Matchmaking Agents project being run at MIT exemplifies an extremely important development for intelligent agents. The project is an experiment in creating a decentralised, fault-tolerant application that handles potentially sensitive information (such as individuals' electronic mail, personal files or lists of particular interests) in a responsible and privacy-protecting fashion, using cryptographic and other techniques. The eventual goal is ubiquitous deployment across the Internet. It is clear that if these agents are deployed successfully then there is much scope for usage in the internal operations of a company.

Personalised news services, such as Individual's FIRST!, can also be utilised as agents for internal processing of information (the WWW site can be found at <http://www.individual.com>). FIRST! is a daily customised news service 'used by corporations to institutionalise current awareness and to create smarter, more dynamic organisations.' Individual, Inc. claims FIRST! is geared towards any workgroup that needs to stay abreast of the daily industry events affecting their business. These groups include sales, marketing, product management, finance, purchasing, advertising, public relations and competitive analysis, among others. News stories are delivered each morning in full text, via fax or electronic mail, or as an enterprise-wide feed for groupware platforms such as Lotus Notes and the WWW. This kind of agent will clearly be important for the internal dissemination of company information.

8.3. Selling

As for purchasing, intelligent agents could play a vital role in marketing and selling processes. Sales leads could be highlighted to a sales team which tie in with the concept of electronic trading opportunities (ETOs). These opportunities are being encouraged by the United Nations as part of its World Trade Point initiative. Companies throughout the world, and in particular in developing countries, are given the opportunity to lodge products and

goods on offer (and purchasing requirements) via Trade Points, Chambers of Commerce Training Councils and so on, which are then passed over the Internet (and in some cases, closed networks) to participating organisations. The BT Electronic Commerce Innovation Centre has recently taken on the role of ETO Associate which entitles the Centre's personnel to receive ETOs regularly from around the world. However, at present these are received via electronic mail and are somewhat 'user-unfriendly' — but intelligent agents will have the opportunity to filter linked organisations and potential sales leads.

9. The future — flexible information working

The concept of flexible information working has been dawning in recent years. However, it remains widely misunderstood by many in the commercial domain. The author's personal understanding and vision of the future lies in utilising information in order to work in an optimally efficient and effective manner, giving the individual (and organisation) far more flexibility in when one actually has to work. This may appear revolutionary to some managers: after all, flexi-hours essentially do not exist outside central and local government offices!

However, intelligent information agents offer the opportunity to transact information flows in a timely manner — that is not to say necessarily accelerated but also decelerated where applicable. As communications across the information superhighway become increasingly efficient there will be less and less need for office-based work to continue. The future of mobile, or flexible, working is evident. An increasing number of organisations are providing employees with not only pagers but mobile phones; not only desktop PCs but laptops; not only paper diaries but palmtop electronic organisers and personal digital assistants.

In his book on the technological future, Negroponte discusses how information is relayed and perceived amongst humans and how computers will have to be developed to operate in a similar fashion (Ref 15). Take the example of a translator of languages: it is often a combination of both lingual and bodily expressions that enhances translation, not just a straight-forward word-for-word representation.

Negroponte goes on to use the well-trained English butler as his best metaphor for a 'human-computer interface'. This agent 'answers the phone, recognises the callers, disturbs you when appropriate, and may even tell a white lie on your behalf!' In addition the ideal butler will be 'well trained in timing, versed in finding the opportune moments and respectful of idiosyncrasies'. The key element here is the aforementioned concept of agents possessing knowledge both about something (a process, a field of interest, a way of doing) and about you in relation to that something (your taste, your inclinations, your acquaintances).

The concept of agents which filter news stories and press releases and book holidays or insurance has already become reality. Another application of agents is in scheduling meetings — for those who have used Lotus Notes or MS Mail and so on, you will appreciate that electronically booking dates and times is far easier to handle with the added functionality of electronic mail. But when an agent (the electronic secretary) can do this on a user's behalf the proverbial nut truly shall be cracked. Furthermore agents can take intelligent steps in scheduling meetings over a period of time to ensure one's diary is not too cramped in the near future. This combination of straightforward knowledge utilisation and logical application is the key to the future of intelligent information agents.

The position of information 'handlers' also needs to be addressed: that is, individuals whose job role is primarily information-based (librarians and information-brokers are perhaps the best examples). As recognised by Hermans, human information searchers usually seek help from information intermediaries such as a librarian (Ref 16). Hermans goes on to state that information workers based in larger companies, or those who require information in a more urgent manner, are likely to use information brokers: 'the Internet offers new opportunities for such intermediary/brokering services. Both human as well as electronic brokers are especially valuable when the number of participants is enormous or when information products are exchanged. Electronic brokers can offer two further opportunities over human ones. Firstly, many brokering services require information processing; electronic versions of these services can offer more sophisticated features at a lower cost than is possible with human labour. Secondly, for delicate negotiations a computer mediator may be more predictable, and hence more trustworthy, than a human one.' The latter argument is clearly one likely to cause debate! However, there is certainly a serious point behind unbiased computers, although their flexibility will of course never match a human's.

10. Conclusion

From governmental developments such as the Information Society Initiative to private sector implementations of information-based strategies, we are seeing the advent of an Information Revolution. The Internet will certainly take its place as the medium of the future for the realisation of true information superhighways. Intelligent information agents will be at the helm of circumnavigating this digital globe in order to feed back essential information to individuals and computer systems alike.

Today there are perhaps just a few obvious applications of such agents such as the delivery of news, booking of holidays and business trips, and scheduling meetings. Each of these require agents to be fed certain amounts of knowledge but also literally to go out onto the networks and hunt down other relevant information. In doing so the agents learn and build a body of knowledge along the way, which is analogous to human development.

The importance of intelligent agents was recently summarised by Hyacinth Nwana, of BT Laboratories:

'... (Intelligent) agents are here to stay, not least because of their diversity, their wide range of applicability and the broad spectrum of companies investing in them. As we move further and further into the information age, any information-based organisation which does not invest in agent technology may be committing commercial hara-
-kiri!' (Ref 17).

This is further emphasised by initiation of a first international conference on autonomous agents, due to take place in California in February 1997, which includes co-sponsorship from British Telecom, Microsoft and Mitsubishi among others (Ref 18). This is just one of some 30 conferences which have tracks on, or relate to, intelligent agents taking place between June 1996 and April 1997. The Distributed Artificial Intelligence (DAI) Unit in the Department of Electronic Engineering at Queen Mary and Westfield College has developed and applied DAI and agent based techniques to real world problems in a wide range of commercial and industrial domains. Applications which have been addressed include: electricity, transportation management, electricity distribution management, particle accelerator control, patient care, concurrent engineering, digital libraries and business process management.

It is possible that agent intermediaries will replace their human counterparts. But what is far more likely is that human intermediaries will have to work in close cooperation with their agents, such that there is an evolution in the tasks performed by both. In the short-term intelligent agents will handle the more mundane tasks of information retrieval and searching whereas humans will undertake the more difficult problem-solving tasks. Eventually this will lead to niche specialisations by humans with the agents undertaking ever more everyday tasks. Further research clearly is important to understand just how this evolution of roles may occur.

Today intelligent agents offer a relatively simple form of functionality, in the main via user input into a form or table which is then unleashed to search. Some of these need to remain running directly from a user's desktop, such as AutoNomy's Agentware, whilst others work in the background simply watching passages of data, such as the Topic Information Agents from Verity Inc. (Refs 19, 20). However, there will be an increasing demand for network-based agents which reside remotely and only report back to users either when requested to do so or when certain relevant information has been gathered. According to Hermans, 'agent-empowered software that is as effective as a research librarian for content search will be available in 1998, and may be expected to be used by a significant number of users near the year 2000.'

Intelligent information agents are the way forward — in fact, the key to the future Information Society.

A.J. Garcia-Sierra
BT Postgraduate Researcher
BT Electronic Commerce Innovation Centre
Cardiff University of Wales
PO Box 907
Cardiff CF1 3YP
UK
Tel: +44 (0)1222 874271
Fax: +44 (0)1222 874301
E-mail: garcia-sierra@cardiff.ac.uk
<http://www.cf.ac.uk/uwcc/masts/ecic/>

References

- [1] Bangemann, M. et al. (1994) *Europe and the Global Information Society: Recommendations to the European Council*, European Union, Brussels.
- [2] Cochrane, P. and R. Smith (1994) 21st century information systems, *British Computer Society Expert Systems 1994 Conference*, Cambridge, pp. 3-17.
- [3] Department of Trade & Industry (1995) *The Information Society Initiative*, <http://www.isi.gov.uk/introisi/introisi.html>.
- [4] House of Commons (1995) House of Commons Debate: the Select Committee on Trade Industry Report on Broadband Networks (the Information Superhighway), London.
- [5] Barnes, J. et al. (1994) *Internet Futures*, BT Laboratories.
- [6] Patrick, J. (1996) Renaissance through the Internet, *6th EC World Congress of EC Users & Expo 1996*, Vancouver.
- [7] Wallis, J. and P. Burden (1995) *Towards a classification-based approach to resource discovery on the Web*, University of Wolverhampton, http://www.scit.wlv.ac.uk/~cm5234/project/agent_info/dataman.html
- [8] Maudlin (1995) Measuring the Web with Lycos, *Third International World-Wide Web Conference*, 11 April, <http://lycos.cs.cmu.edu/lycos-websize.html>
- [9] Wayner, P. (1995) *Agents Unleashed: A Public Domain Look at Agents Technology*, Academic Press, Inc., Chestnut Hill.
- [10] Cochrane, P. and R. Smith (1994) 21st century information systems, *British Computer Society Expert Systems 1994 Conference*, Cambridge, pp. 3-17.

Online Information 96 Proceedings

Page 208

- [11] Durfee, E. et al. (1995) *Intelligent Agent Infrastructures for Supporting Collaborative Work*, <http://ksi.cpsc.ucalgary.ca/courses/547-95/bentley/AGENTS.html>
- [12] Maes, P. (1990) Situated agents can have goals, in P. Maes (Ed.) *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, The MIT Press, London.
- [13] Stephens, K.R. (1995) Behaviour analysis and intelligent agent theory, *CIKM-95 Workshop on Intelligent Information Agents*.
- [14] BT Laboratories (1996) *BT Technology Journal*, to be published October 1996.
- [15] Massachusetts Institute of Technology Media Laboratory, <http://www.mit.media.edu>
- [16] Negroponte, N. (1995) *Being Digital*, Vintage Books, New York.
- [17] Hermans, B. (1996) *Intelligent software agents on the Internet: an inventory of currently offered functionality in the information society and prediction of (near-) future developments*, Tilburg University Masters Thesis, <http://www.hermans.org/agents>.
- [18] Nwana, H. (1996) Software agents: an overview, *Intelligent Systems Research AA&T*, BT Laboratories, Ipswich. <http://www.cs.umbc.edu/agents/papers/ao.ps>.
- [19] The First International Conference on Autonomous Agents, Marina del Rey, California, 5-8 February 1997. <http://www.isi.edu/isd/AA97/info.html>.
- [20] AutoNomy home page: <http://www.agentware.com>.
- [21] Verity, Inc. home page: <http://www.verity.com>.



Intelligent information harvesting architecture: an application to a high school environment

Uri Hanani and Ariel Frank

Bar-Ilan University, Israel

Abstract: In the educational arena, information is conventionally scattered throughout many projects and documents and on many systems. This distribution of data inhibits students and faculty members from searching and accessing information conveniently and efficiently. This research project aims to consolidate the disparate data into one information repository. Known as the KATSIR system, the project is developing and implementing a comprehensive architecture for intelligent information retrieval in open systems. The novelty of this approach is the combination of a new research paradigm in information retrieval, called information harvesting, with a K12-friendly interface. This paradigm enables both teachers and students to gain practical experience in harvesting information both locally and throughout Internet sites in a K12-environment. As part of this research, an innovative information retrieval project was developed. The programme is targeted mainly at the establishment and implementation of a comprehensive Educational Digital Library. This new virtual school library was implemented in the Gilo Comprehensive High School on a local area network that contains more than 150 personal computers with a CD-ROM based system, a high-speed line interface to the Internet, and advanced information science tools. This paper presents the KASTIR system with its various components and capabilities on the Internet as a powerful search and harvesting engine, and its promising contributions to the educational environment.

Keywords: Internet, information retrieval, open systems, intelligent information harvesting, computers in high school, multilingual (Hebrew) support

1. Introduction

Due to the huge volume of information gathered on the Internet and its heterogeneity, search engines were developed to enable users to find their way in the World Wide Web. More specifically, in the educational arena, the information is often scattered throughout many projects and documents, and on many systems. As a result of the recent trend of using more computers in school, faculty members prepare their background materials on the computers, and pupils surf the Web as part of their research and doing homework.

Moreover, the transition from frontal teaching to a teaching environment based on information retrieval and telecomputing poses a new challenge. One can note a 'Tower of Babylon' of information resources and educational work that is scattered throughout separate projects and documents. In other words, this distribution of data inhibits students and faculty members from searching and accessing information conveniently and efficiently.

This problem exists primarily in a mature environment of K12 institutions that are pioneers in the introduction of computers to the educational arena. One such pioneer is the Gilo Comprehensive High School at Jerusalem, Israel.

In the early 1990s, the Israeli Ministry of Education and Culture appointed a committee for telecomputing schools all over the country. In 1992, the committee submitted a proposal for integrating computers into the education process in Israel. Since then, much effort has been made to increase the use of computers at schools as a part of a project known as 'Tomorrow 98'. As part of this strategic plan, ten schools were chosen to represent various models of implementing telecomputing environments throughout this project.

The Jerusalem Gilo Comprehensive high school, as one of these schools, has made an effort to construct a model of a virtual library school based on information retrieval, information science paradigms and telecomputing environments. This advanced new virtual library is implemented on a local area network that includes more than 150 PCs, with a CD-ROM server-based system, and a high-speed line interface (frame-relay) to the Internet. The networked system also supports hypermedia authoring tools, a sophisticated video studio and advanced information science tools such as SQL Engine, HTML and harvesting tools. This existing technical know-how and

capabilities enable the teachers and pupils to take advantage of sophisticated search strategies for locating relevant information from the local networks and on the Internet, and to develop advanced Web home pages, Web hypermedia presentations and useful courseware. The full view of Gilo Comprehensive is partially reflected at the Gilo home page. There, some of multimedia projects can be viewed and retrieved. The URL is <http://www.gilo.jlm.k12.il>.

2. Digital libraries and the virtual school

Apart from the pedagogic innovations introduced through advanced information technologies in the Gilo school model, its management tackled the above-mentioned problem of abundance of dispersed documents and media. It was evident that a new coherent understanding was needed to encounter the dilemma involved in the collection and management of digital materials. Here, as a combination of academic know-how and real life experience, the technology recently termed 'virtual digital library' was applied (Ref 1).

One has to comprehend that the concept of a digital library (DL) in a virtual school expresses the revolution in education that has emerged from information technologies and telecomputing. This can lead one day to an environment that is free from the constraints of time and space. Teachers and students can have interactions without attending specific classes according to a rigid schedule. The connection between teachers and students is achieved by means of telecomputing systems based on communication networks. Moreover, most information resources are not limited to the conventional school or library. Computerised databases and communication networks make access to information possible for anyone from anywhere at all times.

This future (or actually the present environment in our case) invokes some new possibilities (as well as questions) such as:

- the role of the teacher as a mediator instead of being a classical owner of knowledge (Ref 2);
- augmentation of the traditional librarian with machine-oriented 'intelligent agents' (Ref 3);
- new methods of human-computer interactions rich with assistance and guidance (Ref 4, 5).

Based on this technologically advanced environment, many DL projects have been initiated at the Gilo High School, to the benefit of both students and teachers. Using the DL, students are able to consolidate several related projects into one information repository. The system also gives students the opportunity to access Hebrew documents, regardless of format, through a Hebrew interface. Ordinarily, this poses a problem due to different computer operating systems, Hebrew formats and general interoperability problems. The infrastructure constructed can also be enhanced to support a multilingual environment (besides English and Hebrew). Another important property of the DL is its 'K12-friendly' interface that is designed to make high school students comfortable and enthusiastic about conducting searches and viewing their results.

To support all these requirements of the educational environment, an innovative project was initiated to develop and implement a comprehensive architecture for intelligent information retrieval in open systems. The novelty of this approach is the combination of a new research paradigm in information retrieval called information harvesting with a K12-friendly interface. This paradigm enables both teachers and students to gain practical experience in harvesting information both locally (intranet) and throughout Internet sites within the K12 environment. The Gilo project is targeted mainly at the establishment and implementation of a comprehensive educational digital library, or in other words a virtual school library. The educational information retrieval system operating at the school is based on the Harvest system originally developed at Colorado University (Ref 6) and is titled KATSIR — 'K12 Advanced Touring System based on Information Retrieval' ('katsir' in Hebrew means 'harvesting').

3. KATSIR description

We give here a concise and brief (and slightly technical) description of the KATSIR project that is the major theme of this paper. We start by a short survey of the Harvest system that served as one of the starting points of our research.

3.1. The Harvest approach — information gathering and access

The Harvest project was launched at the start of the 1990s at Colorado University (Ref 7). It operates as a server on the Web and is targeted to achieve three main goals:

- (1) an infrastructure architecture that can collect (Harvest) distributed indexed information from the Internet, in an efficient manner, with minimal overload on the networks and communications channels;
- (2) detailed customisation of different sorts of indexes over a wide spectrum of varied databases, heterogeneous schema, information resources and URLs;
- (3) support for local caching and information replication, to provide fast response time and the sharing of computing resources.

The Harvest system consists of six components:

- (1) **Gatherer:** deals with the collection of indexed data from the Internet using local providers in an incre-

Online Information 96 Proceedings

Page 212

mental way, and in sending it from the local cache in a compressed stream. It is based on the Essence subsystem that specialises in different data types formats and their retrieval. Essence deals with summarisation, after scanning the original document, by extracting important information from the original document. In this syntactic phase, Essence can be guided to 'understand' and filter significant knowledge and catalogue it along with the indexing process. In HTML format, the different parts are annotated by such special tags as Header, Title or Bold. The Gatherer can be directed to carry on its branching search from URL to URL, or to stop after several link traversals.

- (2) **Broker:** supports a user query interface on the Harvest host over the collected indexed data. Different brokers can operate in parallel, retrieving information from their own host or from all brokers in the harvest network. The brokers update their indexes in an incremental manner whenever a query is invoked or data is updated. A special Broker HSR (Harvest Server Registry) is responsible for the information about all brokers, Replicators and Caches in the entire system.
- (3) **Index/Search Subsystem:** a general interface to Internet search engines (such as WAIS, Glimpse, Nebula). It supports Boolean search and incremental updates. Glimpse is equipped with an efficient indexing system and interactive queries are used as the standard user query interface.
- (4) **Replicator:** replicates data in a weakly-consistent manner over distributed file systems in the network. The replicated data is kept in the network with mirror copies using suitable protocols.
- (5) **Object cache:** its main aim is to manage the system caching memory, to optimise searching for files and data over the network, and save redundant accesses to retrieve data.
- (6) **Object system:** handles complex types of objects that are kept in the network. It allows the storage and the retrieval of objects from local and remote hosts.

Summing up, the Harvest system is constructed as an open and scaleable architecture using Resource Discovery methods. It is customisable, flexible and adaptive to many applications. More technical information and a list of Harvest sites and uses can be found at Ref 7 and at the URL <http://harvest.cs.colorado.edu/>.

3.2. KATSIR objectives and phases

The objectives of the KATSIR project outlined in the paper are as follows:

- the development of a representative model of referential and abstracts materials (Educational Repository) in the Gilo digital library, with specific emphasis on K12 educational documents;
- using the HARVEST Internet search software developed at the University of Colorado to develop and implement an open architecture for intelligent information retrieval, using a friendly K12 human-computer interface, while applying methodologies of 'information brokers', 'intelligent agents' and harvesting tools;
- evaluating the prototype developed by applying it to an educational repository including the drawing of conclusions, from students and faculty use and feedback, that can be relevant for the entire educational community.

To carry out this pioneering project, several phases were outlined:

- investigating the Harvest system and its tools;
- building the prototype educational information data repository as a Digital Library;
- developing a K12-friendly user interface;
- testing and evaluating the prototype with student and faculty feedback.

Before describing these phases and their specific results, we note that the KATSIR architecture framework main contribution is to enable high school students to feel a sense of accomplishment when accessing information and computer technologies. Students are encouraged to gain skills in investigating information systems and in using practical applications of various computer technologies.

4. KATSIR components

KATSIR is composed of four main processes, as follows:

- (1) the collection (gathering) of initial K12 Internet relevant information and URLs;
- (2) information processing and summarisation;
- (3) the presentation of processed information and its retrieval;
- (4) the analysis and follow-up of system usage.

4.1. Initial K12 collection

The gathering process is activated by applying a Harvest Gatherer to an Internet K12 relevant URL. The initial URLs are provided manually to the Gatherer by information scientists or by system users. Hereafter, the Gatherer follows the HTML links from these URLs and collects the relevant documents using recursive branching.

This automatic branching process can be controlled by the system, either by instructing it to stop after a

specific number of branches or by a filtering algorithm (this filtering process is still in a development phase and evaluation).

In contrast to the Harvest original Gatherer, our improved Gatherer is capable of collecting documents that reside on the local user network (high school LAN). Here, using a special KATSIR user interface, the student or teacher can update the database with relevant material such as multimedia presentations or any other relevant documents. This data entry option allows the inclusion of the document description and storage location in the LAN. Thus, the DL is enriched with many interesting materials from the local school environment.

4.2. Information processing and summarisation

Here, four sub-processes are available:

- (1) Summarisation;
 - (2) Hebrew support;
 - (3) Optimisation and cataloguing;
 - (4) Indexing.
- **Summarisation.** Here the Harvest Essence tool is applied. The original collected document is scanned and several important details (such as Document Title, URL, short description, Key Words and a few lines from the document initial content) are registered by use of an abstracting algorithm. This abstraction process is syntactically oriented. However, as mentioned before Essence can be guided to give a special treatment to HTML tags. It should be noted that in order to save disk storage, the original collected document is not kept in the database, but it can be found using its original URL in the Internet.
 - **Hebrew support.** Native English speakers are usually not aware of the problems caused by the Semite languages (Hebrew, Arabic) in the information retrieval arena. Due to writing and reading from right to left (in contrast to Western languages), special treatment must be given to documents that are written in Hebrew. KATSIR supports Hebrew and is capable of converting the document summary, abstracted by the previous phase (summarisation), to a special format that is compatible with the user interface (which operates either in Windows mode or other GUI systems). This necessitates a set of special Hebrew algorithms that deal with the Hebrew orientation indexing and retrieval parts in KATSIR.
 - **Optimisation and cataloguing.** Due to the KATSIR educational environment, we developed further the properties related to filtering, optimisation and cataloguing of the documents. After the automatic summarisation done by Essence, the information scientists can:
 - delete or add keywords to the document;
 - correct any field or description annotated by the system;
 - decide whenever the document is relevant or not;
 - attach any document to the system browsing outline tree that guides the users in their queries.
 - **Indexing.** The original document summaries are processed to produce indexes keys that serve the search engine. KATSIR uses the Harvest standard search engine, namely Glimpse.

4.3. The presentation of processed information and its retrieval

KATSIR allows two main interaction styles (or access methods): information *retrieval* and *touring* (i.e. browsing). The retrieval of information follows Harvest standard text retrieval mechanisms that were tailored to the KATSIR environment. The second method is one of KATSIR's main innovations. Here, the user can take a tour along a topics tree that represents the various educational entry points for both students and teachers. The system automatically generates an HTML page which displays the topics tree that contains links to the relevant materials.

In order to simplify the system usability for the typical K12 user, the user is advised to make use of Glimpse's full (and complicated) search options by special branching. In parallel, KATSIR also advises the user to conduct a simple structured search by a 'frame text'.

4.4. The analysis and follow-up of system use

All activities in KATSIR are logged and monitored, including users' touring steps. This will allow analysis of system usability and performance, and may suggest further research.

5. KATSIR applications

5.1. System integration and evaluation

A main objective of the KATSIR project direction is its K12 orientation. We consider its usability for a typical K12 user, either faculty or student, to be one of the key issues. So, system integration and evaluation were carefully

Online Information 96 Proceedings

planned to evolve as follows:

- (1) system analysis and users' requirement definition;
- (2) prototyping and pilot;
- (3) system evaluation and test runs;
- (4) further development;
- (5) further evaluation and refinement.

System evaluation and tests were made by special teams of students and faculty called 'leading groups'. These groups were consolidated during a two-year process of intensive work, as part of the Gilo representative model, and they also serve other objectives in the school modeling environment. For example, the Faculty leading group chose the information domains that should be included in the topics tree, and they gave the initial feedback to the documents and URLs that were gathered by the research group. In addition, the leading students participate intensively in the evaluation of the effectiveness of the KATSIR query interface. Without these contributions, KATSIR would still be just an academic tool.

5.2. Applications

Taking advantage of the Gilo school daily involvement in all the above-mentioned phases, the students and faculty members are part and parcel of the KATSIR development. The many users involved and the educational projects that were integrated into the DL were just results of this approach.

To illustrate what was achieved in this pedagogic domain, we list here some of the educational projects that are part of these activities as DL components (see Appendix):

Students' projects

- specials events in Israeli political life and atmosphere;
- information banks such as: stamps, NASA activities, participation in special education activities and work groups;
- designing tutorials for other students in the domains of: HTML, 3Dstudio, multimedia tools, etc.;
- electronic newspaper, updated bi-weekly, related to school life as well as politics, current affairs, etc. This project includes also a reference to two scientific electronic newspapers available in Hebrew.

Information for the faculty

- International projects of K12 teachers and discussion groups;
- multimedia presentations done by colleagues and even students that served as reference materials for classes, in many domains such as:
 - physics;
 - humanities;
 - history sciences;
 - geography;
 - holocaust studies;
 - mathematics, geometry and trigonometry;
 - biology.

Internet section

- HTML tutorials;
- Unix tools and tutorials;
- repository of graphical aids to prepare home pages and converters;
- Q&A material and FAQs.

In Figure 1, we try to illustrate all digital library materials in their natural educational environment. The objective is to represent all school activities as one complete information tree, representing a digital library. This metaphor is user friendly and simplifies the DL touring, while making the browsing and retrieval easy for faculty and students. This ongoing sub-project is evolving as a result of the above-mentioned process of evaluation and refinement. We have a future plan to convert it to three-dimensional space, using VRML tools.

BEST COPY AVAILABLE

Online Information 96 Proceedings

Page 215

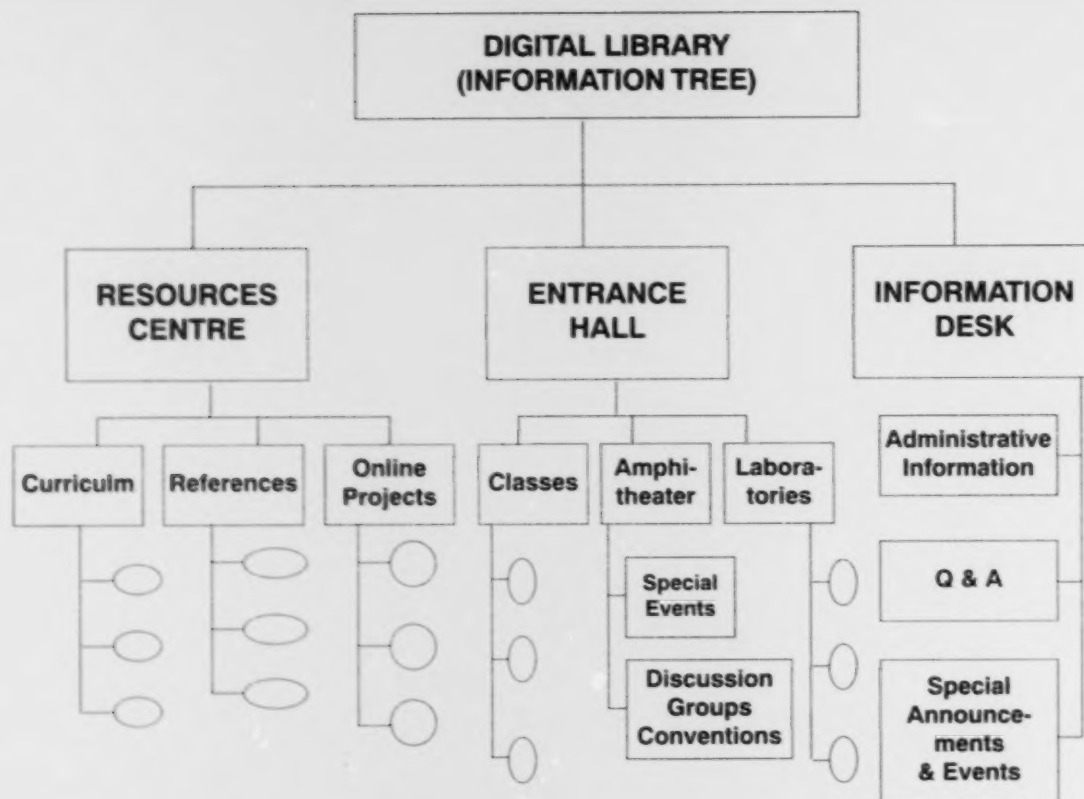


Figure 1: Digital library as a metaphor of K12 environment.

5.3. Local library — intranet

In the analysis study phase we have identified a typical problem that characterises the K12 sector. Not every school can afford an Internet connection, as is possible in the Gilo case. So, an intranet structure (i.e. a local Internet TCP/IP network without immediate access to the Internet) can be installed on the LAN. One essential Intranet tool is a search engine that supports a local DL. Therefore, an SQL engine was tailored to work with KATSIR as an additional tool that handles local digital materials and documents. This tool contains a simple SQL search engine and a topics tree, with a user guided interface, as illustrated (partially) in Figure 2. This local DL can serve as an intermediate repository to the KATSIR DL, and materials can be collected and exported from this repository as from any other Internet repository.



Figure 2: Search engine SQL based digital library repository.

6. Further Developments in KATSIR

KATSIR was initiated as an open infrastructure architecture for intelligent information retrieval and touring. Therefore, it is natural that some innovative developments of KATSIR are under work at the present. We list here some of these developments:

- *Intelligent human-computer interface* that includes a *thesaurus* and consulting facilities, based on the experttext ideas presented at Hanani (Ref 5), Rada (Ref 8) and Chen *et al.* (Ref 9);
- *Developing the Harvest collection process* (the Gatherer) by applying new filtering algorithms based on ideas developed by Shapira *et al.* (Ref 10) while researching user filtering;
- *Developing a three-dimensional human-computer interface* based on VRML and Java applets, using the metaphor of school environment and school layout;
- *More advanced ranking* of search results presented in a friendly way to the K12 users.
- *Applying the KATSIR infrastructure* to regional or even country-based environments (such as urban Jerusalem or Israel), by interconnecting several brokers with a single KATSIR management system. This will allow the application of the caching mechanism of Harvest, with optimisation of search accesses to the Internet, and really arriving at a community virtual school and virtual digital library.
- *Applying a more advanced usability evaluation* other than the user-task model (see Ref 11), and studying advanced productivity tools in the digital library arena (see Ref 12).

7. Conclusions

In this paper we presented the innovative approach of the KATSIR system, based on the Harvest architecture, as an Internet collection system and digital library applied to the educational environment. One major achievement is the cooperation between the academic research team and K12 students and faculty members that allowed the realisation of an educational digital library operating in the field. Our KATSIR environment is an open architecture

that can form a generalised infrastructure to be used successfully by any K12 schools and organisations.

The Internet is going to be the main issue in information science, both as an academic theme and as an application technology. The information gathered at the many (million or more) URLs, home pages and sites will not be available to the user community unless we pay attention to user needs and develop a more adaptable paradigm that will replace the old notion of search engines. Here, the KATSIR project serves as a model as to what that can be done successfully without a large budget and with real life applications. Thus, we have enough background and experience to believe that this research is an important contribution to the information retrieval community of new and beneficial tools and concepts, that can be put to use now and in the near future.

Acknowledgements

This research was funded by the Israeli Ministry of Science and Arts, in the framework of experimentation and demonstration in the domain of information superhighways, digital library and data repositories. The authors are also indebted to I. Miller, KATSIR Chief Knowledge Engineer, and the rest of the KATSIR research group of Bar-Ilan University, Ramat-Gan and the Gilo High School, Jerusalem, Israel.

Dr Uri Hanani
Department of Information and Library Studies, Bar-Ilan University
Ramat-Gan
Israel 52900
Tel: +972 3 5318351
Fax: +972 2 784368, +972 3 5347601
E-mail: hanani@gilo.jim.k12.il

References

- [1] Schatz, B. and H. Chen (1996) Building large-scale digital library, *IEEE Computer*, **29**(5), May 96, 22-26.
- [2] Brewer, A. et al. (1996) The role of intermediary services in emerging digital libraries, *Proceedings of the 1st International Conference on Digital Libraries*, Bethesda, MD, March 1996, pp. 29-35.
- [3] Atkins, D.E. et al. (1996) Towards inquiry-based education through interacting software agents, *IEEE Computer*, **29**(5), 69-76.
- [4] Rao, R. et al. (1995) Rich interaction in the digital library, *Communications of the ACM*, **38**(4), 29-39.
- [5] Hanani, U. (1995) Intelligent hypertext: a frame for automatic search guidance system, *Proceedings of the 19th International Online Information Meeting*, London, pp. 219-229.
- [6] Bowman, C.M. et al. (1994) Scaleable Internet resource discovery: research problems and approaches, *Communications of the ACM*, **37**(8), 98-107.
- [7] Bowman, C.M. et al. (1995) Harvest: a scaleable, customisable discovery and access system, *Dept. Comput. Sci. Univ. Colorado, TR CU-CS-732-94*, March 1995.
- [8] Rada, R. (1991) *Hypertext: From Text to Expertext*, McGraw-Hill Book Company, London.
- [9] Chen, H. et al. (1993) Generating, integrating, and activating thesauri for concepts-based documents retrieval, *IEEE Expert*, 25-34.
- [10] Shapira, B., A. Raveh, P. Shoval and U. Hanani (1996) Hypertext browsing: a new model for information filtering based on user profile and data clustering, *Online & CDROM Review*, **20**(1), 3-10.
- [11] Gentner, D.R. and J. Grudin (1996) Design models for computer-human interface, *IEEE Computer*, **29**(6), 28-35.
- [12] Wiederhold, G. (1996) Digital libraries, value and productivity, *Communications of the ACM*, **38**(4), 85-96.

Appendix

The structure of the DL is somehow a result of the physical layout of the Gilo school infrastructure. The computer network described as follows in Figure 3 and illustrated in Figure 4 may serve also to demonstrate to the reader the computing resources of this project.

GILO HIGH SCHOOL COMPUTING RESOURCES

Configuration Student Labs: Hardware

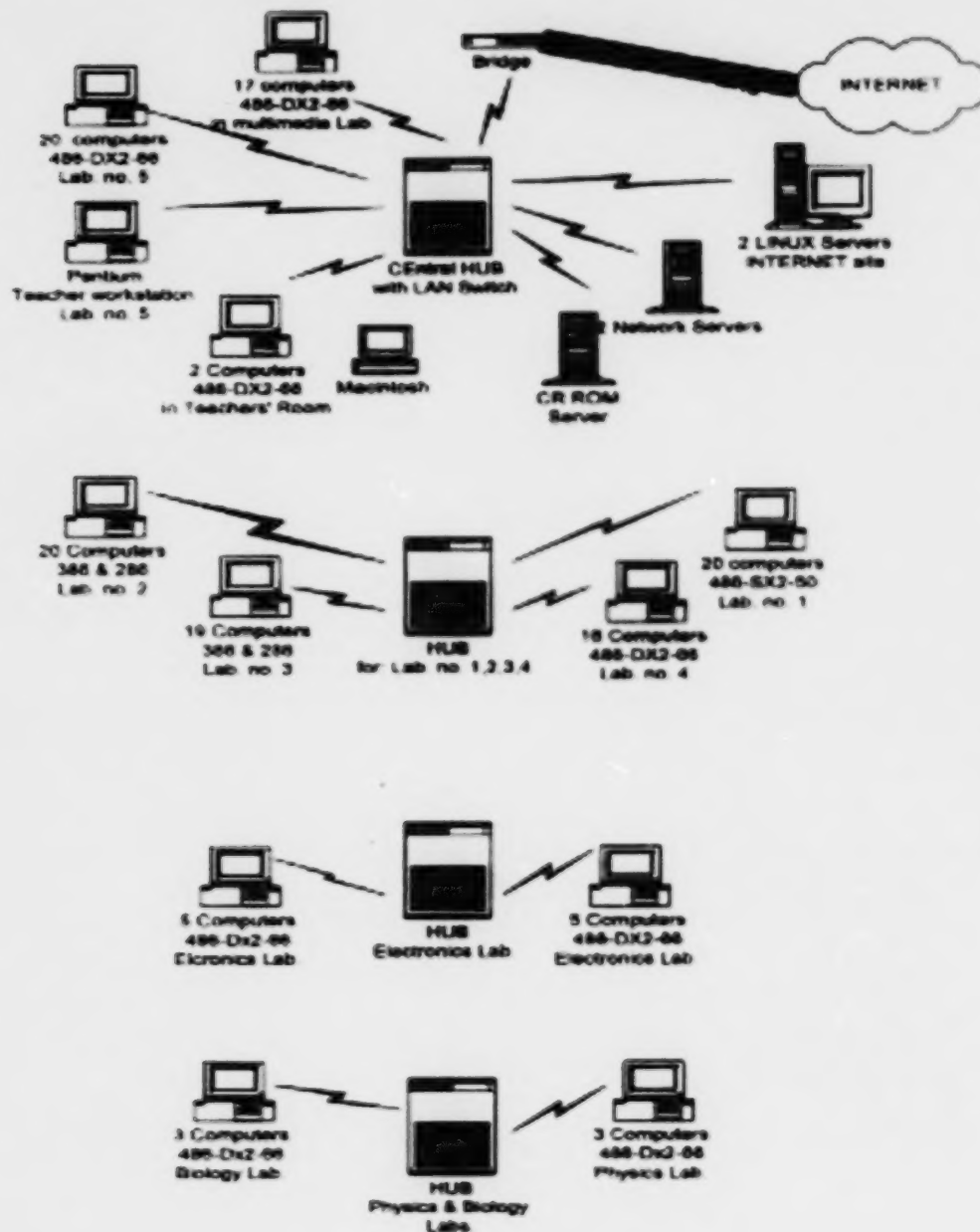
- 1) **Two Novell 3.12 File Servers:** (150+ workstations) serving 7 computer labs, 3 science labs and 10 administrative nodes. All are networked on an optical fiber backbone with a distributed star topology running through a 10BaseT cabling system. Novell 3.12 soon to be updated to version 4.1.
- 2) **Two secondary servers:** a) CD-ROM server working on a cd-rom tower of 8 drives, allowing concurrent sessions on single or multiple drives. b) EDUNETICS server of classroom learning sessions.
- 3) **Two Unix servers:** running on Intel 486DX and Pentium platforms (LINUX). They are connected to a dedicated 64k frame-relay line Internet connection containing: a) World Wide Web server that holds the Gilo High School web site www.gilo.jlm.k12.il.bushwack b) Mail Server (luke). c) "KATSIR" knowledge robot searching selected Internet sites on educational issues. d) four line dial-in modems - soon to be expanded to a multi-line modem Digiboard card.
- 4) **LANNET Intelligent Hub Switchers:** routing data packets on the ten segments of the network. The switcher is capable of: a) supporting combination of 10 and 100 Mb/sec network cards. Now we are running at 10 Mb/sec. b) supporting high bandwidth ATM technology for use in transferring voice and video packets. That is, the infrastructure ready to support video conferencing and multimedia presentations.
- 5) One **Print Server** spooling print jobs to 10 printers.

Software configurations:

- 1) Microsoft Windows and a full suite Microsoft application running from the file servers.
- 2) Microsoft, Novell, Linux, Borland application and software development products are running on individual stations with the aim of having them available network-wide in the future. Also Visual Basic, C++, and Turbo Pascal.
- 3) Multimedia and education authoring systems: ACTION, Astound, PowerPoint.
- 4) Two e-mail systems servers - one on the Unix server running SENDMAIL for staff and teacher use, the second on the Novell server running Mercury that connects students to internet e-mail.
- 5) World Wide Web authoring tools - HTML utilities that allows over 100 students to have home pages on the Internet.

Figure 3: Description of the Gilo High School computing resources.

Schematic Layout of the Gilo Computers Network



Remark: All the HUBs are interconnected to one network with fiber optics.

Figure 4: Schematic layout of the Gilo High School computer network.



New searching technologies and interfaces

Michael L. Weiner and Peter F. Rusch

Manning & Napier Information Services, USA

Abstract: The increasing demand for information has resulted in an increasing number of duties for the corporate library staff. With the advent of the library without walls, librarians and information specialists are tasked increasingly to provide a variety of digital library services beyond the traditional intermediated search. Searches of high economic value remain the purview of the skilled information specialist who has both subject and information skills to apply to the searches required. Preliminary screening searches, competitive intelligence searches and searches of lower economic value are now performed by the end-user of the information satisfying his or her own needs. These end-users may search infrequently and they usually have fewer formal information skills than their professional information colleagues. A new range of databases and search engines with natural language interfaces are available to assist the occasional user perform high-quality searches. The processes embedded in these systems provide an ease of use more closely related to the research skills of the end-user. The techniques of database preparation, query formulation and search execution using the DR-LINK interface will be described.

1. Introduction

For many years research in information retrieval has followed several paths leading to development of both experimental and commercial systems. Historically, the straightforward word- and phrase-indexed search systems were commercially implemented earliest. These systems took advantage of a confluence of events in technology and database development. Improvements in computer technology through the 1960s brought both faster central processors and larger random-access disk storage, each of which benefited both information retrieval and database development.

The improving computer technologies were applied to database collection, processing and production usually as an adjunct to some printed publication process. Although initially small, many of these databases today contain millions of records and billions of bytes. Some organisations followed dual development paths, working on both database development and information retrieval development. By the early 1970s the commercial database industry was generally divided into database development, practised by a group of producers (usually publishers), and commercial information retrieval from these databases, offered by the growing online services of the day.

Databases adopted a variety of old and new techniques in creating their content. Standards for bibliographic data evolved and were used. Subject specific indexing schemes were applied and augmented as the databases grew. Information retrieval depended solely upon the words and phrases entered by the database producer. Typically, database production was an extension of manual production processes that were assisted by computer technologies in the capture and preparation of magnetic distribution media. Accordingly, database content was usually brief and was (and still is) subject to human error. Thus, concept searching was performed using word, or phrase, searching that required allowance for any number of misspellings and word variants. Inference and nuance were lacking. Controlled vocabulary indexing relieved some of the problems encountered but led to the need for information specialists to search effectively the specialised vocabularies that developed. Further, these controlled vocabularies evolved as the subject matter itself evolved and changed. As time passed and databases grew larger, even controlled vocabularies lost some of their value since they were not consistently applied throughout the time span of a database.

Commercial, online information retrieval systems of today are often characterised as 'Boolean' or statistical/probabilistic retrieval systems. Perhaps a more apt designation is 'search term based' (or STB) systems. The systems are designed to retrieve search terms, whether they are words or phrases. Databases are prepared for search by deriving search terms and presenting these to the user. Such information retrieval systems are truly WYSIWYG systems in the sense that even misspellings appearing in the databases become search terms and retrieving these items depends on anticipating the misspelling. Users are generally required to specify the databases where they believe the information resides. They give a detailed query composed of search terms that are words and phrases hoped to describe the query in the context of the databases' coverage. Results are presented as a group of 'hits' that must be sorted manually for relevance to the query posed.

2. Improvements to information retrieval

As the number and size of databases grew so did the information retrieval problems associated with them. New kinds of databases appeared. Commercial information retrieval services grew, added more databases and adapted to the changing needs of both databases and users. More full-text was available as the result of so much text being captured from word processing. No longer were secondary (abstracting and indexing) databases the only means of information retrieval. Search and delivery of the complete document became expected.

At the same time, computer technology and telecommunications were becoming ubiquitous so that the number of people equipped for information retrieval grew enormously. With relatively few information specialists to handle an increasing number of requests, new computer-based methods were needed to assist this growing number of information seekers. Results from the information retrieval research community were refined and implemented to advance information retrieval. Many of these techniques are incorporated into DR-LINK (Document Retrieval through Linguist Knowledge) that was created to help automate the process that research librarians use to convert information needs, as stated by users, into the information retrieval process.

3. The DR-LINK search system

DR-LINK resulted from participation in the US Government initiative called Tipster that was sponsored by the Advanced Research Project Agency (ARPA). This was an effort of significant proportion to stretch the limits of information retrieval. The test queries used to evaluate the responses of competing information retrieval systems included temporal and consequential aspects requiring linguistic knowledge. Indeed, the idea was to expand the capabilities of information retrieval to process not only queries formulated in natural language but to search and retrieve natural language 'documents'. The results have been impressive, and substantial improvements in precision and recall can be demonstrated.

To illustrate the expectations of the project look at an example Tipster query on the topic, 'Black Monday' (Ref 1):

A relevant document will contain at least one reason why US stock markets experienced a huge drop on 19 October 1987, losses of equity so large that markets were said to have crashed (the Dow, for example, lost 508 points on that day alone); the date of the crash has become known as 'Black Monday.' A preferable document would contain a detailed analysis of the crash. The best document would link analysis of events to actions taken or recommendations made by federal authorities or the stock markets to prevent further crashes. *Not* relevant are reports which simply reference, without analysis, 'Black Monday,' such as anniversary stories generated by the press around every October 19th.

The DR-LINK search system was designed and developed to handle queries of the kind given above. It has been demonstrated to do so successfully. DR-LINK actually contains several independent retrieval technologies that are blended to create a new information retrieval process. Central to this new process are the concepts of psycholinguistics. Scientists in the field of psycholinguistics have determined that humans process text on at least six known levels: morphological; lexical; semantic; syntactic; discourse; and pragmatic. The **discourse** level refers to the all of the meaning we derive beyond the word and sentence level such as reference to information provided previously in an article. It also refers to the difference between definite information related to the past, and prediction and expectations in the future. The **pragmatic** level refers to that which we know outside of the text of an article. References to general nouns imply the specifics: for example, 'third world' implies the countries that are known to comprise that grouping.

Working with this model of six levels of text processing, the value of applying them to information retrieval can be illustrated. For example, if we encounter 'White House' in a document about the US Government or the US President, we know at a pragmatic level in this context it refers to a branch of the United States government, physically located in Washington, DC in the USA in North America. We also know that it is *not* a European nation, nor a Third World country, nor is it South America.

Natural language processing (or NLP) systems such as DR-LINK use this kind of ancillary information in a way that permits them to become part of the search process. Both the databases at the time of loading and the search query at the time of execution are processed using these relationships to build a richer environment for information retrieval.

A search query about future, evaluative comments on the outcome of a specific pending event is a very different request from one about historical results. Many queries contain this temporal and consequential component. Statistical/probabilistic online information retrieval systems rarely deal with these degrees of subtlety. To the extent it is available, it may be part of a controlled vocabulary indexing system. Thus, effective use of statistical/probabilistic systems may require extra efforts by trained information retrieval specialists to formulate the query iteratively and review the results in detail.

A search system that interprets and expands on a user's natural language query automatically can expand the access to information without the help of a trained information retrieval specialist (Ref 13). This works to the benefit of the trained information specialist in two ways. Ideally, it frees the trained searcher to focus on the most difficult and strategically vital searches. Even for the information specialist, the improved results of a search permit more time for analysis and interpretation. Overall usage of internal and external information in the organisation increases.

4. Specific applications of DR-LINK

In the area of searching for prior art for patents, this type of linguistically informed search capability could change fundamentally the manner in which patents are prosecuted. If the patent offices accessed such a powerful search system, one that could overcome some of the complexities and obfuscations of language, what patent applicant could afford to be without a comparable, comprehensive search?

With a complete database collection it would be possible to search not only newspapers, magazines, abstracts and bibliographic data but also entire reference books, specialised encyclopedias and complete doctoral dissertations. All of these represent significant amounts of non-patent prior art that has been impractical to search previously.

DR-LINK, with its large collection of databases containing computer software prior art, is able to locate significant prior art that could be used to challenge patent claims. An example query of this nature is:

'I would like information about searching for textual, graphical, audio or video information in a multi-media software environment that combines these attributes into a single, interactive, related program for entertainment or education.'

The top ten articles retrieved from the collection of databases containing computer software prior art are on point and should be reviewed seriously with respect to the disclosures made in them (Refs 2-11).

Competitive intelligence is similar in the desire to obtain *all* prior information. In competitive intelligence the objective is to gather all information about a competitor and its products and services. Using a natural language profile that searches a variety of document types from many localities it is possible to alert an organisation to any competitor's new or rumoured products. Such a system can return results of reasonably relevant information, interpreted, expanded and filtered to match what was needed rather than results that merely matched lexically.

Likewise, a market survey search is best done with a variety of sources to get a broad view of some market opportunity. An example of such a query is:

'I would like information about point of sale devices, linked to computer systems, that scan bar code displays and then compare the product code to certain data to generate discount coupons for the customer automatically at the point of sale such as a supermarket checkout counter cash register.'

The top five ranked articles to appear from this query have titles that are relevant to the topic of point of sale coupons in supermarkets (Refs 14-17). The rank order of the documents is that given among the references cited and illustrates the benefits of retrieval over a broad collection of diverse databases. More than a dozen databases were searched, ranging from full-text, primary sources such as newspapers and research reports to secondary abstract and index.

5. Conclusion

The benefits of the DR-LINK search system permit natural language queries to be input by trained information specialists or novice users. For the novice, the query can be formulated as an expression of interest to a research librarian who would convert the information need into the information retrieval process. For the experienced searcher, the process allows more time to be given to analysis of results. Searching goes beyond the tradition search term-based system, allowing for automatic term expansion and the removal of ambiguities of search terms. Consequences of events and temporal references may be part of the query. Results are presented in a ranked order presenting the most relevant items to be scanned first.

Acknowledgements

The authors want to thank the staff of TextWise Co. who helped with some of the examples presented.

Peter F. Rusch
355 Verano Drive
Los Altos
CA 94022
USA
Tel/fax: +1 (415) 941 8120

References

- [1] Liddy, E. (1994) Information retrieval via natural language processing or an intelligent digital librarian, *ASIS Mid-Year Conference*.

- [2] Mackay, W. and G. Davenport (1989) Virtual video editing in interactive multimedia applications, *Comm. of the ACM*, pp. 13.
- [3] Stevens, S. (1989) Intelligent interactive video simulation of a code inspection, *Comm. of the ACM*, pp. 15.
- [4] Halasz, F. (1988) Reflections on NoteCards: seven issues for the next generation of hypermedia systems, *Comm. of the ACM*, pp. 29.
- [5] Barbour, B. (1990) Enhancing the BBC Domesday Videodisc as a learning environment for New Zealand secondary schools, *Brit. J. Educ. Technology*, 4.
- [6] Warwick, M. (1993) Survey of telecommunications in Business (10), *Financial Times*, 12 July 1993.
- [7] Sonsino, S. (1993) Survey of information and communications technology, *Financial Times*, 18 April 1993.
- [8] Schnase, J.L., J.J. Leggett, D.L. Hicks and R.L. Szabo (1993) Semantic data modelling of hypermedia associations, *ACM Transactions on Info. Systems*, pp. 1
- [9] Gibbs, ., D. Tsichritzis, E. Casais, O. Nierstrasz and X. Pintado (1990) Class management for software communities, *Object-Oriented Programming, Comm. of the ACM*, pp. 25.
- [10] Davis, J.I. (1992) File 1 — a computer & information technologies platform, *Computer Underground Digest*, pp. 18.
- [11] Chen, C.C. (1990) DATAPLEX: an access to heterogeneous distributed databases, *Comm. of the ACM*, pp. 15.
- [12] Weiner, M.L. and E.D. Liddy (1995) Intelligent text processing and intelligent tradecraft, *The Journal of AGSI*, July, 60-67.
- [13] Feldman, S. (1996) *DR-LINK, DIALOG and TARGET: A Comparative Study* (in publication) (a complimentary copy of this article may be obtained by writing the authors).
- [14] Gellene, D. (1994) Check it out supermarkets track top patrons, reward them with custom deals, *LA Times*, 4 August.
- [15] Anonymous (1988) Interactive POS video yields instant results, *Chain Store Age Executive*, pp. 1.
- [16] Coleman, L. (1988) 'Smart Card', coupon eater targeted to grocery retailers, *Marketing News*, pp. 1.
- [17] Bradshaw, D. (1994) Technology: the end of the queue — automated supermarket checkouts will cut waiting time, *Financial Times*, 13 July 1994.
- [18] Tomkins, R. (1994) Management (marketing and advertising): time to cut it out — US manufacturers are questioning the value of money-off coupons, *Financial Times*, 18 May 1994.



Automated patent searching in the EPO: from online searching to document delivery

Annemie Nuyts and Charles Jonckheere

European Patent Office, The Netherlands

Abstract: *The European Patent Office (EPO) has recently implemented the last part of its ambitious automation project aimed at creating an automated search environment for approximately 1200 EPO patent search examiners. The examiners now have at their disposal an integrated set of tools offering a full range of functionalities from online searching, via full text browsing, to document delivery on the workstation screen or a nearby departmental printer.*

Online searching, via a common command language, is carried out in a great number of bibliographic and full-text databases on the EPO in-house host service or in commercial databases (on STN, DIALOG, QUESTEL etc.). Specially developed 'dual-mode' viewer software enables the examiner easily to browse full-text documents (in character mode) and view patent images (in facsimile mode). Browsing and viewing, as online searching, are interactive operations with very short response times. The recently installed electronic document server, finally, gives access to the EPO's complete search collection comprising some 25 MIO documents (250 MIO pages compressed according to CCITT group 4.T6), totalling approximately 14 terabytes of storage space. The storage technology is based on the use of magnetic cartridges placed in robots, which offer not an interactive online service but rather a 'near-line' service with maximum 15 minutes response time.

Keywords: online searching, common command language, viewer, browsing, document delivery, European Patent Office, EPO, full-text searching, dual-mode, near-line, EPOQUE, BNS

1. Introduction

The European Patent Organisation is an inter-governmental organisation set up pursuant to the European Patent Convention which entered into force in 1977. At present, 18 Member States have ratified the Convention (15 of the European Community as well as Liechtenstein, Monaco and Switzerland).

The European Patent Convention establishes an organisation to implement a single procedure for the searching and examination of European patent applications. The executive body of the EPO is the European Patent Office which has its head office in Munich and branch offices in The Hague, Berlin and Vienna. The European Patent Office employs approximately 3700 persons of whom some 2000 are patent examiners (search and substantive examination).

The EPO's main line of business, the delivery of patents, involves the handling and processing of vast amounts of information. Automation is the key answer to dealing with the ever increasing volume of information. The Office is already an extensive user of automation and is expanding its use considerably over the coming years.

Traditionally the EPO patent examiner, for performing a patentability search, has made use of a systematically classified paper collection comprising some 25 MIO patent documents. Over the past years, however, automated tools were installed progressively for supporting electronic documentation and search activities. The present paper gives an overview of the service as now available to close to 2000 end-users, among which the approximately 1200 search examiners are very intensive users.

2. The problem of automating a patent office

One of the important targets in automation projects in the worlds' major patent offices is the field of patent documentation and prior art searching. Indeed, traditional searching methods based on classified paper documents have reached the limits of their capacity. The sheer volume of published documents, currently growing at the rate of around 800,000 new items a year, inevitably leads to longer search times and higher costs, since examiners have more and more information to contend with. Subdividing groups of documents and making

classification more exact can offer only a partial remedy to the problem.

Moreover, more and more we are seeing an ever-increasing number of cross-disciplinary inventions extending over several technical fields at once. Electronic games are one example of this trend. Inventions of this kind mean that more search groups of documents have to be consulted and make classification a more complex task than in the past.

Finally, patent applicants now have access to a wide range of online information systems which are used with great skill. In the majority of cases the ability and the technical know-how to make an accurate assessment of the state of the art before filing the application is at the applicant's disposal. The credibility of patent offices would be called directly into question if their attachment to traditional, i.e. paper methods rendered them unable to obtain results as good as or better than those achieved by applicants themselves.

Patent documents are not just articles containing information: they are highly technical and pose several specific problems: for example they include a large number of drawings, essential to the understanding of their content, and their text relies on a sophisticated and sometimes highly esoteric language and is difficult to read. When designing automated search systems these aspects have to be taken into account. A system based solely on bibliographic, abstract or full-text searching will always be of limited use — text has to be supplemented by graphics. It must be necessary to navigate through the document, moving from a drawing to the relevant part of the text — and vice versa, jumping to another section of the text relating to the same graphical image — without having to scroll through the complete document in order to find the next interesting passage, compare one diagram with another and so forth.

Electronic searching also has to be interactive and follow an iterative rather than a sequential logic. As well as being expensive to set up, automated systems based on sequential consultation of a large number of documents do not and cannot meet the real needs of patent examiners.

Finally, it has to be remembered that searching methods vary in subtle and significant ways from one technical field to another and within a technical field, even from one patent application to another. This is already the case with the traditional paper tools: it is all the more so with their electronic equivalents. The available data, the user-interface and the way to work with the automated tools therefore have to foresee to a large extent the different ways of working that are dependent on the technical field involved (Ref 1).

3. The three basic elements of the EPO solution

The aim of the EPO's automation programme, therefore, was to build an electronic information system whose scope equals or exceeds that of traditional paper documentation. Reviewing the basic requirements of the search examiner and matching them to an electronic search leads to the conclusion that the fundamental process to be supported comprises the following three basic steps (Ref 2):

- identifying the set of documents which are potentially relevant to the patent application at hand (typically 200 documents);
- eliminating irrelevant documents and selecting relevant documents (typically 20 documents);
- in-depth study of selected documents.

These three steps quite often are repeated interactively several times before reaching the final goal, typically consisting of four or five pertinent documents cited in the search report.

The *EPOQUE suite of applications* is the cornerstone of the automated patent search environment at the EPO. In order to support the three steps described above it consists of three major parts:

- an *online search and retrieval tool*, with growing emphasis on the user's added value input;
- an *online viewer part* providing display and browsing of the full documents for one member per patent family — including the first page abstracts and images — in dual mode with ASCII text and facsimile drawings, in the selection/elimination process;
- the electronic equivalence of the entire paper document collection, providing copies (paper and electronic facsimile) of the complete original documents for in-depth study before citation in the search report.

These three parts will be detailed in the following paragraphs.

4. The EPOQUE search and retrieval tool

The EPOQUE host service was installed on the EPO mainframe computer in 1989 (Ref 3) and has been further developed ever since. It enables the search examiners to interrogate a great number of databases from their workstations.

In a common environment, with a sophisticated user interface, interactive interrogation is possible of databases from three different sources:

- internal (databases loaded on the EPO mainframe);
- external (databases loaded on external commercial hosts);
- personal (databases created by the examiners themselves).

The aim is to have these three retrieval parts permanently available for use with one single standardised way

Online Information 96 Proceedings

of interrogating them (common command language). The user interface that permits this includes the functions Internal, External, Personal, Preparation and macros, Download, Function Keys, Language and Data Conversion. A simple click with the mouse or combination of keyboard strokes as well as the 'drag & drop' possibility will activate whatever option the user wants to select.

4.1. Main EPOQUE retrieval functionalities

All the internally loaded databases used for searching and retrieval in the EPO have been standardised as much as possible. For instance, in the whole set of databases the patent, priority or application numbers, dates, document types or classification symbols are always written in the same format with the same field name. This allows easier cluster searching and direct cross-file between databases.

The examiner must be able to connect automatically to external databases loaded on commercial hosts, and to interrogate these using the same EPOQUE command language (language conversion) and using standard data format for patent numbers (data conversion), as well as standard field names (tag conversion).

The personal database tool is designed to allow the creation of specific databases on the users' hard disks, enabling the addition of personal codes, remarks or keywords to the documentation in the examiners' own specific search field. This tool also uses a subset of the EPOQUE language and the same database structures.

Important search aids have been created such as the so-called 'standard functions' which are in fact super macros of commonly occurring questions. They can be used without use of the command language. The standard functions are permanently available for selection on the screen, ready for use, wherever else the user may or may not be connected to.

The workstation software includes offline 'preparations', (search strategies) the ability for users to program macros and associate them to function keys, and the ability to share these preparations. The continuous creation of the LOG files on the users' workstation is seen as a major feature in the daily use of EPOQUE. Whatever happens in the dialogue between internal, external or personal sessions, everything is written to a file. This feature not only permits going back and forth in the whole online session while being online but it also allows, in append or overwrite mode, the entirety of sessions executed in the past, at any moment in time, to be consulted.

4.2. Internal databases

The internal databases comprise in the first place those which are produced by the EPO itself. Among these the EPODOC database, which contains the bibliographic data of the EPO's patent documentation as well as abstracts and titles, plays a key role. Many more databases, such as for example EUREG (European Patent Register) and ECLA (European classification scheme), provide valuable information to the search examiner in his or her daily work. EPOS is one of the databases created by the users themselves. EPOS is a collection of synonyms or expressions related to a certain technical concept and organised per technical field.

In the second place come the databases produced by Trilateral partners, such as PAJ (Patents Abstracts of Japan) which gives access to the English language abstracts of patent applications filed with the Japanese Patent Office. From the US Patent and Trademark Office the EPO receives bibliographic data and full texts of US patents, as well as from various other databases such as UCLA (US classification scheme), while the incoming US abstracts are included in EPODOC.

In the area of full-text patent databases the EPO is building up a set of databases covering the so-called PCT-minimum documentation published after 1970 (one patent per patent family). Of this collection more than 2.5 MIO documents are already available in text coded form as they were produced via an automated printing process. The missing 1.3 MIO documents are obtained via an OCR conversion process (see Section 5.2).

Thirdly, the EPO has loaded on its internal host some databases delivered by external producers, currently comprising:

- WPI (World Patent Index produced by DERWENT);
- INSPEC (Information Services for Physics, Electronics and Computing produced by IEE The Institute of Electrical Engineers);
- TDB (Technical Disclosure Bulletin produced by IBM).

More databases in this category will follow in the near future, among which a series of full-text journals published by ELSEVIER is worth mentioning.

4.3. FIRST PAGE — first step to access the full document

In the patent world, almost all patents now have a standardised first page containing bibliographic data, title, abstract and a 'pertinent' image.

The FIRST PAGE software was able to reconstruct on the workstation the first page of the patent for display on the screen, from the abstracts and bibliographic data available in EPODOC and WPI and the 'clipped image' of the front page. FIRST PAGE became available in 1992 but with the recent introduction of the EPOQUE VIEWER system the first page information is available in this new model at the same level as the complete text and all drawings of the full documents, thereby reconstructing the complete application in dual mode, as detailed in the next section.

5. The EPOQUE VIEWER — a browser and navigation tool

During 1995 the EPOQUE VIEWER came into service. It provides display of the full documents including the first page abstracts and images, for one member per patent family in the selection/elimination process. This online dual mode access uses the ASCII form of the full text of the applications (contained in full-text databases) together with facsimile images of the drawing pages (contained in image databases), thereby enabling examiners to eliminate documents from a result list and/or tag documents for further, deeper study either on paper or on screen via the BNS (see Section 6.1).

The success of this patent information system is dependent on whether the viewer part is capable of solving the users' needs when looking at complete documents on screen, which is always a painful task to perform. In this respect every examiner is able to combine the single relevant items of information contained in a full document that he or she wants to see. This combination may always be the same (e.g. drawings and description of the drawings) but it often changes case-by-case.

It is also important to enhance the users' efficiency compared to paper searches by giving added value to full document browsing. Functionalities which cannot be dealt with in conventional paper searches are focusing on different levels of highlighting, real navigation inside a document and, as close as possible, automatic positioning at a needed passage. Allowing the user to mark relevant areas inside the documents is an essential part of this added value navigation (Ref 4).

5.1. EPOQUE VIEWER functionalities

The user may choose which data to display in patent and non-patent literature documents: abstract and bibliographic data, first page image, the full patent document, patent claims and/or drawing pages — all this is made highly flexible. All combinations of all the different pieces of information are possible.

The search for the unique family member present in the EPOQUE VIEWER is for the system to take care off: a list of PNs given by the user is sorted and de-duplicated, and the corresponding documents (only one per family of identical priorities) present in the VIEWER are searched.

The display of the requested data (browsing through full text and drawing images) is done by total interaction of the user with the system. The workstation takes care of the necessary requests for the selected pieces of text and/or drawing image pages which are stored centrally.

Additional string searching is offered through a dialogue box 'HIGHLIGHT ... string of text', repeated automatically when more text comes in. The retrieved strings of text are highlighted and it is possible to jump from one highlight to the next.

The tagging functionality helps out in the way users eliminate or keep documents from a result list. Some examiners go through the whole set quickly just to get a rough idea of the possible pertinence of the documents and then come back to look at a number of pre-selected documents again. Others study every document more thoroughly from the very beginning and therefore need much more detail because the pertinence of the documents looked at may influence the progress of the search.

In order to add value — by the examiners themselves — to the navigation aspects of EPOQUE VIEWER, marking is possible in the full text, especially in the description and in the claims, by adding the marking word immediately before the sentence in the record. This feature is completely controlled by the end-users who are able to see immediately on the screen the markers they have added, modified or deleted.

5.2. Coverage of the EPOQUE VIEWER document collections

The EPOQUE VIEWER collections will cover as much as possible the information needed by the examiners to carry out the selection/elimination process of the search procedure. For the documents published after 1970, EPOQUE VIEWER provides access systematically to the full text and images of the so-called minimum-PCT documents. Per patent family, only one document is available: approximately 4 million unique documents. The same family member provides the text, the drawing images and the FPAGE information (abstracts from EPODOC, WPI, PAJ and clipped image of the first page). Those documents that have text available in ASCII form are loaded immediately. The others are captured by OCR (Optical Character Recognition) techniques of the corresponding BACON facsimile copies. A three-year contract for OCRing the documents needed to fill up the backfile is ongoing.

6. The EPOQUE-BNS viewer — the access to the original document

The third main part of the EPOQUE suite is the BNS application (BACON Numerical Service). This electronic equivalent of the entire numerical collection provides copies (paper and electronic facsimile) of the complete documents for in-depth study before citation in the search report. It is expected that from the combination of an EPOQUE search and VIEWER elimination and selection process, on average only 20 documents will remain in the final 'hitlist'. The BNS provides the examiner with a copy of those 20 documents (in display on screen or in paper form).

6.1. The BNS service

The BNS service which became available early 1996 is not an interactive online service: it works in so-called near-line mode offering an average response time to the end-user of approximately 3— minutes (the contractually agreed maximum response time is 15 minutes for 98% of the requests). Today the BNS is delivering copies of about 16,000 documents per day. This number includes near-line access (around 3000 documents per day) as well as additional services performed as batch processes, such as the production of extra copies of documents cited in the search reports for mailing to the patent applicants, copies for documentation purposes (8000 documents per day), copies for National Offices of the member states of the EPO, and so on.

The data included in the BNS currently consist of almost 25 million documents in numerical order containing about 250 million scanned page-images compressed according to CCITT Group 4 T.6 representing a total volume of around 14 terabytes. The yearly increase is 3–4% of the total volume (approximately 800,000 documents). The data are stored on double length, double density cartridges of 800 MB each, randomly placed in three robots, each containing 6000 cartridges. An active backup duplicates this entire collection in three additional robots (in a separate room for security reasons), with all the tapes organised in reverse order for faster access.

6.2. The BNS usage

As the third tool in the process of online retrieval and display the BNS will deliver the 'original' document to the examiner within maximum 15 minutes, thus allowing the users to keep concentrated on the subject matter which was being analysed during the selection phase.

The use of the BNS workstation interface is straightforward. A working list is created from the tagged PNs in the EPOQUE Viewer and a request sent out to the BNS server. Because of the near-line access mode, several working lists may be issued. Once the BNS server has retrieved the needed documents from the robots, they are copied to a buffer and from there downloaded to the workstations.

The user can consult an 'overview window' and get information about the status of the issued requests. When a request is in the 'ready' status the documents can be displayed in their original form, by browsing from page to page or switching between different parts (description, figures, claims bibliographic information, amendments, etc.) within a patent publication. Zooming facilities and double window display (for comparison between two different pages on one screen) are available.

A major component of the BNS system is the BPS, the BACON Printing Service. Tagged pages, whole documents with tagged pages or specific parts of documents can be printed either on decentralised departmental laser printers or on centralised printers. The priority for the different print batches depends on the number of pages requested to be printed. For detailed in-depth study before the citation in the search report, examiners print approximately 10 documents or parts of documents. The BNS also allows overlays, i.e. electronic labels with information from databases and integration in the printing job of documents from other sources like search reports.

7. Conclusion

The EPOQUE suite of applications represents an important achievement in the area of online patent searching at the EPO, bringing together all major patent and some non-patent literature databases in the world in one standard system. The workstation software is setting new standards of efficiency and comfort for online end-users. With the combination of EPOQUE RETRIEVAL, VIEWER and BNS, the EPO can justifiably claim to have one of the world's most advanced set of tools for automated patent searching.

In order to ensure that the potential of the new tool is fully exploited, the EPO launched two important projects involving all search examiners: the 'learning process' with the goal of refining and perfecting the new working methods, and the 'bottom-up approach' working as a continuous programme to match the different tools and their contents to the needs of the examiners working in different technical fields. The important challenge, indeed, facing the EPO is the management of the evolution of the culture — the creation of new methods of work, their implementation and the adjustments that are necessary at the individual level: the most difficult of all to change but the most vital to maintaining an efficient and high quality patent search.

C. Jonckheere
European Patent Office
Patentlaan 2
PB 5818
2280 HV Rijswijk (ZH)
The Netherlands
Tel: +31 70 3402564
Fax: +31 70 3403320

References

- [1] Michel, J. and K. Douglas (1994) The future of search in the European patent system, *World Patent Information*, **16**(4), 193-197.
- [2] Van Thielen, J. (1994) Search tools at the European Patent Office: a new approach, *World Patent Information*, **16**(3), 137-139.
- [3] Jonckheere, C. (1990) EPOQUE (EPO Query service): the inhouse host computer of the European Patent Office, *World Patent Information*, **12**, 155-157.
- [4] Nuyts, A. (1995) A new dimension in patent indexing and searching — the use of navigation tools, *Proceedings of the 1995 International Chemical Information Conference*, Nimes, France, 23-25 October, pp. 78-88.



Providing database services in a nationwide research organisation — coexistence of traditional information services and a modern CD-ROM/online hybrid solution

Benjamin F. Bowman

Information Retrieval Services, Max-Planck Institute of Biochemistry, Germany

Abstract: For the past two decades the central Information Retrieval Services of the Max Planck Society have been providing database searches for scientists in Max Planck Institutes and Research Groups throughout Germany. As a supplement to traditional search services offered by professional intermediaries, we have recently fostered the introduction of a new information system that can be accessed directly from all sites. Ovid Technologies was chosen as partner for this project. Their common user interface offers database-specific search tools for the expert as well as for the novice user. Further advantages are the use of the TCP/IP protocol allowing immediate access from a number of different computer platforms, the client/server architecture and the adherence to the Z39.50 standard, thus providing interconnectivity with other library servers. The possibility of using the Ovid Local Loader to install internal databases under the same user interface as external databases was of particular interest to our Institutes. The advantages of a CD-ROM interface in combination with the performance of an online system ('CD-ROM/online hybrid') has made this new service very popular among our scientists. Of critical importance for the quality and success of the project was the active participation of experienced information specialists during the planning, testing, installation and final production stages. We continue to play an active part in supporting the end-users with workshops, documentation, advice and a help-desk for database specific questions. The Information Retrieval Services are now confronted not only with a variety of new responsibilities but also with many new opportunities for helping the users to gain the information they seek.

Keywords: information retrieval services, end-users, intermediaries, CD-ROM/online hybrids, database searching

1. Background

The Max Planck Society (Max-Planck-Gesellschaft, MPG) encompasses approximately 100 Institutes and Research Groups which are located all over Germany (there is also one Institute in The Netherlands and one in Italy). These pursue basic research in the natural as well as in the social sciences, including such diverse fields as biology, chemistry, medicine, psychiatry, astronomy, physics, mathematics, science history, educational research, international law and many others. There are approximately 4000 scientists plus an equal number of students doing research in Institutes ranging in size from only a handful of co-workers to a staff of over 1000 at a single site. All Max Planck Institutes and Research Groups are assigned to either the 'Biological-Medical Section', the 'Chemical-Physical-Technical Section' or the 'Humanities Section', depending upon their main scientific activities. Funding of the Society is provided mainly by the German government but also by grants from various sources.

Obviously, researchers active in the forefront of science need to have the best possible access to scientific information. The libraries of the Institutes have always played a vital role in providing information resources, both through their own holdings and more recently by making CD-ROM databases available to their patrons. However, the increasing amount of information being produced, the diversity of the media being used, the complexities of installing and managing information systems, and the simultaneous reduction of library budgets and staff have made it very difficult, especially for a smaller library, to continue providing a high level of competency and service.

When bibliographic databases first started becoming available in the early 1970s, scientists of the MPG immediately took an interest in them. However, because of the amount of time, effort and money necessary to make efficient use of these new information sources, they were hardly being used. Groups of specialists were

soon established at two of the larger Max Planck Institutes (the MPI for Biochemistry near Munich and the MPI for Solid-State Physics in Stuttgart) in order to provide mediated database searches for the Institutes of their respective Sections. The specialists in these newly established Information Retrieval Services had previously been active scientists in the fields in which they are now carrying out searches. This proved to be of considerable value for the quality of the results and for the general acceptance of the service (Ref 1).

For over 20 years both groups have been actively serving the Institutes, always in close collaboration with the libraries. The 'traditional' services include online searches in all major databases (mainly through DIMDI and STN), as well as setting up, processing and distributing almost 1000 weekly literature reports (using SciSearch tapes from ISI). There has always been a very close relationship between the Information Retrieval Services and the Institutes they serve. Each Institute is visited on a regular basis (usually every 18 months) by the responsible member of the group. Extensive interviews are held with all recipients of the weekly reports and with any other members of the scientific staff in order to optimise current search strategies, to initiate new searches and to discuss other topics related to information management (Ref 2).

2. 'Managing the end-user'

Although the online hosts worked hard at improving their user interfaces, end-user searching did not become popular with our scientists until some of the more important databases became available on CD-ROM. With this new media and its more attractive interface, searching seemed much easier and the users were very eager to try it. The increasing number of databases available on the Internet has now led to the situation that almost all our scientists have experience in searching databases and mediated searching has become the exception rather than the rule. Fascinated by their new possibilities, the users do not always realise that they might not be getting a true picture of the available information simply because they are not sufficiently familiar with the contents or structure of the database they are using (Ref 3).

In recent years there has been a growing tendency in our Institutes to purchase and install CD-ROM databases locally. This was of course very welcome to the users, very prestigious for the libraries and very profitable for the producers. Although we fully support the concept that researchers must have direct access to databases, we were concerned about two important aspects: first, which criteria were being applied when choosing the databases or retrieval software and second, would direct access be interpreted as being a full replacement for professional searches? We wanted to ensure that, regardless of their choice, the scientists in our Institutes would be able to receive the most reliable search results possible, with or without the help of an intermediary (Ref 4).

We started taking a closer look at some of the information products being offered to end-users. Although we restricted our investigation to only a few important databases in our field, we soon encountered a considerable number of different versions. There were at least a half-dozen CD-ROM versions of the MEDLINE database with differences in the contents, in the retrieval capabilities, in the years covered, in the networking possibilities and so on. Current Contents were available on Diskette or on CD-ROM, one week at a time or a 52 week rolling file, with or without abstracts, single or combined editions. Often the selection of one of these versions by the users was due to coincidence, or according to the price tag, rather than an accurate knowledge of the differences. In fact most of the scientists were not even aware of the existence of other versions of the databases they were using.

The situation was especially obvious with MEDLINE, a database that we know very well through our extensive online usage and that was already being used in various CD-ROM versions in many of our Institutes. The very sophisticated indexing methods of this database helps the user to find highly relevant articles that he might otherwise have missed but only if he knows which terms are available and how to use them. Since the average user is not willing to spend much time reading manuals or help texts, the retrieval software and the user interface must efficiently guide him through the database and help him to formulate a successful search strategy. In our comparison we found that those MEDLINE CD-ROMs that performed best in this respect were not necessarily those which our scientists knew about or were already using.

From our own searching experience we knew that reliable search results depend greatly on the choice of the most appropriate database(s) for the particular question, knowing the unique structures and indexing policies of the selected database(s) and having a powerful retrieval software available that allows for the construction of efficient and database specific search strategies in each database. Given the general lack of broad searching experience in the Institutes on the one hand, and their independence in financial decisions on the other, we felt obliged (in the interest of the scientists) to become actively involved in the selection process for end-user oriented products or systems. However, despite our offers to inform and advise the Institutes before they invested more time, money and manpower, we frequently were only consulted if problems occurred after a system had already been bought and installed. In some cases the decisions in the Institutes had been influenced by inexperienced end-users who simply were not aware of the diversity of products available and the significance of their differences, and therefore had not thought it necessary to get more advice.

We realised that the only effective way to support end-user searching in our Institutes was to offer access to a system which had so many obvious advantages that it would be preferred to any other. If the Institutes could decide on a common system this would, of course, also mean that the Information Retrieval Services would be in a better position to offer support and training. The options at the time were either reaching agreements with online hosts and producers that would allow direct access to their databases (possibly at special rates), or setting

up a centralised service within the Max Planck Society based on a selection of CD-ROM databases. Taking into consideration that the decisions in the Institutes would be strongly influenced by the user interface (rather than the underlying retrieval capabilities), it was clear that most end-users would probably prefer using a CD-ROM based solution.

3. The Max-Planck/Ovid Database Server Project

The CD-ROM user interface that impressed us most was being offered by CDPlus (the name has since been changed to Ovid Technologies, henceforth referred to as Ovid). Their approach was distinctively different from that of other vendors. While Ovid had developed a very innovative common user interface it obviously had also recognised the importance of database-specific features and had not sacrificed them for the sake of simplicity. The use of these features makes considerable difference in the quality of a search. They are, however, often difficult to comprehend and their correct application is frequently a problem, even for experienced searchers. Although other vendors may have kept the features available for use by experts who know where to look for them and how to use them, Ovid has successfully integrated them into easy-to-use search tools that the end-user can apply without the necessity of having prior knowledge of their existence or meaning.

We had tested Ovid's CD-ROM versions of MEDLINE and Current Contents and in both cases we felt that the database-specific structures had been incorporated exceptionally well, and that an end-user could easily learn how to use them correctly. Ovid's user interface allows searches to be entered in various ways according to the individual level of expertise. Tools such as natural language mapping to controlled vocabulary, integrated graphical thesauri, database-specific search limits and simultaneous on-screen user support are not only extremely useful for the novice but are also welcome aids for experts. The end-user is unobtrusively guided towards a deeper understanding of how the database-specific tools and the powerful command syntax can be employed to achieve better search results (Ref 5).

After learning more about the number of databases available under Ovid and, in particular, about the very interesting multi-platform network solutions they were offering, it was decided to consider Ovid for the installation of a central database service for the Max Planck Society. Besides their unique user interface, several further advantages were soon identified. These included the use of the TCP/IP protocol which would allow immediate access from a number of different platforms (DOS, Windows, Macintosh, and VT-100 terminals), the client/server architecture with its inherent division of labour and reduction of network traffic, and the adherence to the Z39.50 standard allowing interconnectivity to many other library servers (Ref 6).

The databases and the Ovid software may be installed on a Unix-based server without the necessity of having multiple CD-ROM drives or 'juke boxes'; CD-ROMs are then merely used as transport media for database updates. In fact, databases offered under the same Ovid Database Selection Menu can be physically distributed over several different servers at different locations on different computer platforms and installed on either hard disks or CD-ROMs, without the end-user noticing any significant differences to a local CD-ROM or hard disk installation. Of particular interest to several of our institutes was the fact that with the aid of an Ovid Local Loader it would be possible for their scientists to search databases produced internally under the same common user interface as the commercial databases being offered on the server.

It was vital to the success of the project that the Institutes themselves be convinced of the quality, reliability and prospects of this new service and thus be willing to pay their share of the costs. It was therefore decided to install the Unix version of the Ovid software on a server located in Göttingen at the GWDG (Gesellschaft für Wissenschaftliche Datenverarbeitung), a computer centre of the Max Planck Society, load a number of databases that would be of interest to a large portion of the Institutes and then provide unrestricted access from all sites for a seven-month trial period before reaching a final decision.

The trial period began in April 1995 with the complete MEDLINE database (1966 to the present) and Current Contents/all editions (1995, cumulative). During the last three months of the trial the BIOSIS Previews database was added. We would have liked to include more databases, in particular some that would be of greater interest to the Institutes of the other two Sections of the Society, but this proved to be difficult to negotiate. Coming to an agreement with Ovid was not the problem but getting acceptable pricing from the database producers was a different matter. The major issue was: should the Max Planck Society (comparable in size to an average German University) be considered as a single site or as 100 individual sites? Regrettably, not all producers were willing to offer a concurrent-user based pricing scheme.

The trial agreement with Ovid included a series of introductory workshops which were held at several of our larger sites throughout Germany. This provided a good opportunity for the Information Retrieval Services to present database specific information to the prospective users that was not readily available elsewhere, but nevertheless important for them to know, in order to select the most appropriate databases and to develop effective search strategies. Many of those attending these workshops had assumed they already knew enough about the databases and were only interested in learning how to access this new system. Afterwards, almost all participants stated that they were surprised by many of the important aspects they had learned about the databases, and that learning how to access the system and how to use the Ovid interface was quite simple in comparison. Thus, in addition to introducing new searching capabilities for end-users the workshops served to underline the fact that database specific knowledge is an essential factor and that this expertise is available to the user through the Information Retrieval Services.

At the beginning of the trial period, access was only possible via a telnet session using a character-based

interface which had nevertheless the same basic layout and similar functionality as a graphic interface. Interest in the service was immediate, and within the first few weeks several hundred users had applied for passwords and were performing searches. It was important for the success of the project that the Ovid Windows Client became available to our users at least during the last three months of the trial. For those scientists working with Windows, this provided a very convincing alternative to local CD-ROM holdings of the same databases. The prospect of soon having the same graphical interface and improved functionality available for the Macintosh and other platforms led to a very favourable reception of the project.

The decisive questions were now: how should the future costs be spread among the Institutes? How much would the individual Institutes be prepared to spend for such a service? Would the total amount received be enough to continue with the present databases? Was there a willingness to pay for additional databases? Various models for the distribution of the costs have been proposed. The yearly costs to the Max Planck Society depend mainly on the number of users who are allowed access to the databases: the currently adopted internal cost-distribution scheme, however, is based on the amount of usage of each database by each Institute (which of course is much more complicated to determine). We would be very interested to learn how other consortia are handling this issue. In our opinion one of the most significant advantages of flat-rate pricing is that it encourages higher usage of valuable information resources, which should be the major objective in the installation of such database services.

In November 1995 the final agreements were signed and the transition from trial period to full production phase took place without interruption. Currently (June 1996) there are approximately 1400 active user IDs and the server is being accessed from 87 Institutes or Research Groups. Nearly 7000 database searches are being carried out per month. Comments on the quality and performance of the new service have been very positive. With the introduction of the Ovid Web Gateway an additional platform-independent point of access has recently become available which has the same basic functionality, plus some very useful new Web-specific features.

Ovid has helped us to establish a modern database service for the Max Planck Society that has all the advantages of CD-ROM databases (for example user-friendly interface, fixed fees) but without many of the disadvantages of actually running CD-ROMs in a network. In addition we have the advanced performance capabilities that are more typical of an online system — a true CD-ROM/online hybrid (Refs 7-9). Although other vendors (e.g. Silver Platter, Knight-Ridder) are now also offering interesting hybrid solutions, we feel they have not yet reached the same level of performance and quality, especially in their implementation of database-specific features.

We consider the current status to be only the initial phase in the development of adequate electronic information resources for end-user searching in the Max Planck Society. Compared to the present situation in most German universities we have made significant progress, but compared to top-ranking US or British universities we still have a long way to go. Access to full-text sources, links to document delivery services, links to library OPACS and the inclusion of further databases are topics that need to be addressed soon. Other models of end-user searching are also currently being tested. A recent fixed-fee agreement between STN and the MPG allows unlimited access to the INSPEC database. A further possibility would be to access databases being offered by online vendors which adhere to the Z39.50 standard directly from the Ovid client or Ovid Web Gateway.

4. An enhanced role for the information professional

The Information Retrieval Services of the Max Planck Society have played a major role in initiating, planning, testing and introducing the new central database services. We will continue to play an active part in supporting the end-users with workshops, documentation, advice and a help-desk for database specific questions. This will help to keep us in close contact with the scientists in our Institutes and ensure that they know where to turn if they need more detailed information or if they prefer to have a search done for them by an experienced professional searcher.

We feel that it is important for our scientists to have *both* possibilities: expert searches through intermediaries and direct access to a selection of databases. Depending on the situation, the user can then choose whatever he feels is the most appropriate method. Our observations have shown us that in the majority of our Institutes, end-user-searching has been welcomed as an important *additional* option and is not seen as a substitute for the traditional services. Despite the rapidly increasing number of searches being carried out by the end-users themselves, there has been only a slight decrease in the number of search requests to the intermediaries. This (momentary?) reduction is quite convenient because it allows us to spend more time with the more demanding searches and with providing support to end-users. The Information Retrieval Services are now confronted not only with a variety of new responsibilities but also with many new opportunities for helping the users to gain the information they seek.

The participation of experienced intermediaries in the decision-making process will continue to be of utmost importance. Care should be taken that the available expertise is not 'sacrificed' for the establishment of services that leave the users only one option, namely to rely on their own search results using systems that were chosen for them by well-meaning but perhaps less-experienced persons.

'Will intermediaries be "disintermediated" or will this mean an enhancement to their role?'

The answer to this question (as posed by the organisers of this conference) lies mainly in the hands of the intermediaries themselves. Only if they are willing to accept the challenges of the new electronic information environment and to participate actively in its further development can they hope to continue having an influence

on the quality of the services available to the users. It is precisely their knowledge and experience in dealing with users and databases which is most urgently needed for making the right decisions, now and in the future.

Dr Benjamin F. Bowman
Information Retrieval Services for the Biological-Medical Section of the Max-Planck-Society
Max-Planck Institute of Biochemistry
Am Klopferspitz 18a
D-82152 Martinsried (near Munich)
Germany
Tel.: +49 (0) 89 8578 3832
Fax: +49 (0) 89 8578 3833
E-mail: bowman@biochem.mpg.de

References

- [1] Urbas, W. (1984) Informationsvermittlung in der Max Planck Gesellschaft — Organisationsformen und Perspektiven, *Nachrichten für Dokumentation*, **35**(3), 124–129.
- [2] Nordwig, A. et al. (1984) Informationsvermittlung in der Biologisch-Medizinischen Sektion der Max Planck Gesellschaft, *Nachrichten für Dokumentation*, **35**(4/5), 199–204.
- [3] Bowman, B.F. (1989) Selbstrecherchieren — Pro und Kontra. Erfahrungen mit Wissenschaftlern und mit Literaturdatenbanken, *Nachrichten für Dokumentation*, **40**, 77–82.
- [4] Bowman, B.F. (1994) Konzepte zur Qualitätssicherung — aus der Praxis einer Informationsvermittlungsstelle, *Cogito*, **10**(3), 30–32.
- [5] Jacso, P. (1994) OVID/MEDLINE — a gourmet choice, *Database*, **17**(1), 31–38.
- [6] Brunelle, B.S. (1995) Platform- and media-independence in the electronic research environment, in *Online Information 95: 19th International Online Information Meeting Proceedings*, London, 5–7 December 1995, Learned Information Europe, Oxford, pp. 209–217.
- [7] Vogt, S. (1992) CD-ROM and online — competitors or compatriots? An online host's point of view, in *Online Information 92: 16th International Online Information Meeting Proceedings*, London, December 1992, Learned Information Europe, Oxford, pp. 33–45.
- [8] Pack, T. (1996) CD-ROM/online hybrids, *Online*, **20**(2), 70–74.
- [9] Lescher, J.F. (1996) DIALOG on the Net, KR Science Base, *Online*, **20**(3), 38–50.



Electronic publishing and collection development, a subscription agent's view

Philip Wallas

EBSCO Information Services, USA

Abstract: Trends in publishing, advances in technology and pressures on library budgets have combined to put libraries and publishers at odds with each other. Research libraries expect broad, easy access to electronic information, greater convenience and faster delivery but at reduced cost. Publishers are exploring new channels for distributing their content while seeking to maintain control of their intellectual property, minimise risk and increase income. Subscription agents have an important role in resolving these tensions.

Substituting access to digital information for ownership of materials introduces new uncertainties concerning delivery systems, costs and collection stability. Electronic publishing asks both publishers and libraries to take on new roles which may be handled better by subscription agents. Business issues such as managing access are obstacles to wider use of electronic publishing. The demands of password maintenance and authorisation are similar to the headaches of ordering, renewing and paying individual subscriptions. IP address authentication and user ID/password access control have predominated; Access Consolidation is suggested as an integrated approach.

Keywords: electronic publishing, collection development, research libraries, subscription agents

1. Differing needs, different expectations

Trends in publishing, advances in technology and pressures on library budgets have combined to put libraries and publishers at odds. The number of titles published continues to grow. Delivery systems have increased the velocity of information flow. Researchers' expectations for information systems have been characterised by the aphorism 'the least they will accept is the best they have ever seen.' Libraries' expectations for electronic publishing include broader access, greater convenience, faster delivery and, it is hoped, reduced cost. Publishers, while exploring new channels for distributing their content, seek to maintain control of their intellectual property, protect their subscription revenue and find new ways to raise income. Subscription agents have an important role in resolving the resulting tensions.

2. Questions to frame the discussion

In our market economies, aren't we used to getting what we pay for? That is to say, don't we expect to pay more for more functionality? From transportation to food to clothing to housing, we pay more for increased speed, quality, style, convenience and security.

Do librarians really expect to get more information electronically that is easier to search and distribute, and more timely, for less than they paid for print?

In recent years, aren't we used to getting more and more technology for our money, whether it is in handheld calculators, long distance calling, personal computers or television sets?

Do publishers really expect to charge more money for electronic versions of their content when their production and distribution costs are greatly reduced by the use of new technologies?

Another current trend word is disintermediation, the tendency for businesses and consumers to go directly to manufacturers for services and products instead of working through distributors and agents.

When everything is available for direct access on the World Wide Web, do subscription agents really think they have a future?

In the next few minutes, let me describe what one subscription agent overhears when librarians and publishers talk about what they hope electronic publishing will mean for them, as well as what they fear it may mean.

3. A research library perspective

Substituting access to digital information for ownership of materials offers libraries new opportunities for broad distribution and more comprehensive research, but at the same time introduces new uncertainties concerning delivery systems, costs, access control and collection stability.

Libraries seek electronic access agreements with generous terms to allow their user community flexibility in using information. Ideally, library users should have unhindered access to search, view, print and download information for their own use, for electronic reserve desks, for course packs and for inter-library loan. No obstacles such as passwords or the need to use particular computer systems should interfere with research. Material from many sources should be seamlessly available for review, collected by discipline and topic. The library staff should not have to administer access codes or install or upgrade special software. Researchers away from the library should be able to access the same information they can search when they are at the library. Walk-in users should be accommodated without difficulty. Electronic systems should provide detailed data on usage, enhancing the ability to make sound collection development decisions. Finally, costs should be lower than existing print subscriptions.

Today's reality is quite different from this ideal. Electronic access to any particular information may be ephemeral. You have all navigated the shifting currents of the Web, where information appears and disappears unpredictably, like a sandbar. With competing technology solutions in the market today it is difficult to decide which solutions to implement when there is no certainty which standards will prevail. Relying on electronic access begs a number of questions about collection stability and access control:

If you subscribe this year to an electronic title and then let your subscription lapse next year, will you have access to the year you 'bought' or will there be a new hole in your collection?

By subscribing this year, do you get access to earlier years of data or do you pay extra for them?

Who will set up and maintain the multiple accounts, passwords and software configurations that publisher-controlled access to multiple servers will require?

As a subscription agent, we have been approached by librarians who value our long-term relationship with them and who need these questions answered. Together we are exploring how our existing electronic publishing business might evolve to meet their needs.

4. A publisher perspective

For publishers, new technologies promise to shorten the time between submission of manuscripts and publication as well as to reduce distribution costs. By providing electronic access, publishers hope to reach new markets. On demand publishing, solely electronic journals, electronic document delivery and e-mail responses to articles offer new possibilities for enriching the information exchange and making it more interactive. Usage tracking and encryption technologies may make it possible to block unauthorised use of content and to capture additional revenue from legitimate copying.

Primary publisher goals include: building reputation for their titles, editors and contributors; maintaining control of content and preserving the predictable revenue flow provided by subscriptions. Electronic publishing might add revenue from delivery of individual articles or from royalties earned for author abstracts. Broader, faster access may increase visibility through wider coverage in indexing and abstracting services. By hosting information on their own servers, publishers can have direct relationships with their customers, collecting significant marketing and usage information which should aid planning for new titles.

Compared to this ideal picture, today's reality is a series of experiments in which publishers are learning just how far outside their traditional business electronic publishing activities are. In each experiment, one or a few publishers maintain a server and provide access to content under a variety of restrictions, using many competing delivery technologies, collecting small amounts of additional revenue while incurring significant costs to develop and maintain software. The more open the access, the more possible that it will cannibalise print subscriptions. The more restricted the access, the more burden is placed on researchers to remember passwords or administer authorisation, and on publishers to manage ordering, access and billing.

Making individual document delivery easier may lead to the atomisation of journals, a kind of entropy in which journals disintegrate into what has been called a 'soup' of articles. Tying electronic subscriptions to continuation of print subscriptions creates a whole layer of account management: print only, electronic only, both, none of the above. Finally, selling and supporting electronic access turns out to be a demanding business requiring a new set of skills. In effect publishers add printer, post office and 24 hours a day, 7 days a week technology consultant to their existing role.

As a subscription agent, we have been approached by publishers who value our long-term relationship with them and who do not want to take on the burden of these additional roles. Together we are exploring how our existing electronic publishing business might evolve to meet their needs.

5. An integrated electronic publishing environment: one agent's perspective

There are a number of established ways to deliver journal information electronically, including simple ASCII representation, bitmapped images and portable document types such as PDF files. A variety of search engines are available for finding desired data. While additional progress will certainly come in both search and display technology, business and legal issues — not technical problems — are now the major obstacle to wider use of electronic subscriptions.

In an environment where each publisher maintains a server for its content, the demands of password maintenance and authorisation are similar to the headaches of ordering, renewing and paying individual subscriptions. Providing seamless access to information and managing access control can be seen as extensions of the traditional role of the subscription agent. Four sometimes conflicting goals for managing access have been described by libraries:

- users should not have to remember user IDs and passwords for each service;
- all users at an institution should have access to some databases;
- some users at an institution should have access to databases not available to others;
- access should be available away from the library.

5.1. Two methods for access control

Two primary approaches are currently used to manage electronic access: user IDs with passwords and IP address authentication. User ID and password systems require the user to enter an ID and related password before access is allowed. Because this method identifies the user, not any particular location or computer, access can be available from many locations outside an institution: e.g. office, home, hotel room or dormitory. In addition, different passwords can have different levels of access, allowing for example staff and faculty to have access privileges not available to patrons or students. Balancing this broadly distributed access is the need to issue, maintain and have users remember IDs and passwords for each system they use.

IP authentication relies on the nature of the underlying Internet protocols which make it possible for a host system to know the Internet address of each user. With this method, the host system maintains a register of the Internet addresses authorised to access a particular database and only allows access from those addresses. No one has to assign or remember passwords. Because of the way IP addresses work, it is easy to authorise access for all the computers at an institution. But since this approach identifies a particular computer, not a person, it is not easy to allow different classes of access for different people. Dial-up and remote access are also difficult to authorise since the IP address to be used may not be known in advance, or it may change dynamically with each session, as with most dial-up Internet service providers. Finally, it is possible for hackers to 'spoof' or falsify IP addresses.

A hybrid of these two methods might allow IP authentication to a library community for a general collection of databases, but require a password for access to restricted databases, or for access away from the permitted IP domain.

6. A new approach: access consolidation

A new approach, access consolidation by the subscription agent, allows libraries the ease of use they seek while giving the publishers the control they need. With the agent's system managing contact with the various publishers' servers, administrative friction in retrieving information, handling billing and managing access is absorbed by the agent. At the library end, a cooperative authentication approach provides both flexibility and control.

In cooperative authentication, the institution adds a link from its Web server to the EBSCOhost online service. When this link is selected, the institution authorises access by asking for a user ID already known to the user (student ID, employee number or library card number). The institution's system verifies the user, then makes the connection to EBSCOhost. By relying on an existing identification system for authorisation, this method allows remote access, dial-up access and different levels of access without requiring any additional user ID maintenance. At verification, different categories of users can be connected to different EBSCOhost user profiles, with different access privileges. For walk-up users without an institutional relationship, public terminals in the library could be set up to use IP authentication for access.

Let me outline some of the different levels of access consolidation the agent might provide:

- *Password consolidation*: publishers allow the agent to store the user's passwords and perform access authentication for one or more journals on different publishers' own servers;
- *Publisher server full-text access consolidation*: publishers allow an intelligent link from the agent's online system to their own servers so citations located on the agent's server can retrieve full text located on the publishers' servers, with authentication and billing handled by the agent;
- *Agent's server full-text access*: publisher allows complete journals to be stored on the agent's server for search and retrieval, with authentication and billing handled by the agent's server

By listening to librarians and publishers, adopting new technologies and participating in standards organisations, agents have created time and money-saving efficiencies for libraries and publishers. With more than 50 years experience of consolidating ordering and billing of serial information, EBSCO is well positioned to play the newly expanded intermediary role of access consolidation. By combining traditional functions with access control, electronic search, and online and paper document delivery, the agent can bring the ideals of both libraries and publishers a bit closer to reality. We are encouraged by the way our customers and publisher partners are responding to our proposed role in electronic publishing.

Philip Wallas
Manager, Rights & Permissions
EBSCO Publishing
10 Estes Street
Ipswich
MA 01938
USA
Tel: +1 (508) 356 6500
Fax: +1 (508) 356 6565
E-mail: pwallas@epnet.com



The Internet and local online databases: a *novum organum* for twenty-first-century library science

Terry Webb

University of Hawaii, USA

Abstract: *The Internet and World Wide Web will become the forum for introducing and retrieving the most current research in all fields. This technology will also be the means for making public vast accumulations of information and knowledge that, although valuable, will never find their way into conventional print publication. For librarians, who in the past two decades have grown from technological recalcitrants to information innovators, Internet/WWW technology has revolutionary implications. It stands to make librarians the direct intermediaries between library users and the producers of new knowledge, and amounts to a new method for library service — a novum organum for library science.*

Keywords: online databases, libraries, Internet, World Wide Web

1. Introduction

When I began my library career over twenty years ago, in the early stages of library automation, librarians were being bullied by vendors and programmers who did not fully understand the mission and processes of libraries. Rampant computerisation pulled us in its tow toward technologies we did not comprehend. It is no wonder we were technophobic. Furthermore, we were hobbled by our own rigid procedures and customs; we saw automation only as a way to perform faster and, we thought, more economically the same bibliographic tasks we had always performed. Only gradually did the bright idea begin to dawn on us that information technologies were creating possibilities of doing things we had never before considered.

Over the following two decades, the library community overcame its technological intimidation. Today, librarians are up-to-speed with information technologies and have become champions of acquiring access to those technologies for their users. We have followed a learning curve composed of a succession of technological dreams and implementations that have included massive bibliographic databases of shared cataloguing, automated circulation systems, online catalogues, CD-ROM servers and document delivery utilities.

The realisation that these shining breakthroughs were not the end goals, but only evolutionary phases, temporary resting places and even misplaced hopes should have provoked in librarians revolutionary ideas and concepts. They should have made us re-evaluate the foundations of our profession. But through all the technologies, for all the talk of new models for librarians, despite all the quibbling over what now to call ourselves — librarians, information specialists, or even cybrarians — have we really changed?

2. Format management

The most recent technological dream to fade is the so-called virtual library. We have learned from several noble attempts that the fully electronic library cannot be created overnight. The reasons are now quite clear (Ref 1):

- (a) it will take a generation for people to get used to exclusively virtual resources;
- (b) it will take that long to accumulate a body of electronic literature sufficient to sustain in-depth multi-disciplinary research;
- (c) the disparate systems that must talk to each other to link digital resources into fully virtual libraries are not yet interoperable, not ready for 'metadata';
- (d) publishers will not accept any threat — real or perceived — to their copyrights.

For at least the next decade or two, then, libraries will continue to select from an assortment of information formats the most appropriate one to meet a specific information need. This strategy, which I call 'format management', is becoming an important part of librarianship. Primarily it involves selection of paper, CD-ROM or online as the appropriate format for an information resource based on the nature of the information itself and on the type of usage it will receive. For example, paper is still desirable for certain types of resources because it

requires low maintenance, is relatively inexpensive and does not entail sophisticated machines or complicated searching strategies. Paper resources are best suited for background and abstract readings, aesthetic works and texts not meant for data extraction.

CD-ROM, though it behaves like an online resource for extracting data, is actually offline. It is not continuously updated, as online is, and so is best suited for less dynamic information, for information of some importance to users, but not enough to justify a subscription to an expensive online database to get it, and for searching database backfiles offline before incurring online expenses to retrieve the most current information.

Online, of course, is the best technology for extracting data from the highly dynamic information resources that are essential to modern research. But for the most part, even the online databases are less current than they appear to be because they obtain much of their content from print publications: that is, they are by-products of the conventional processes of scholarly and scientific publishing which greatly delays the appearance of research. By the time a new piece of research appears in print, or in its alternate electronic format as an entry in a commercial database, it has already made the rounds of the experts and been criticised, revised and perhaps even replicated. It has also already been worked over by editors, referees and typesetters. Even in very dynamic databases we see research that was conducted much earlier. This is because commercial databases are built on the infrastructure of publishers, editors and marketing departments. These online resources are based on research that was already, or is about to be, published in print sources.

The new online methods, which combine publishing and communications technologies, offer researchers the means to share their findings much earlier, to a much larger circle of peers, even to the entire scholarly community at a point much closer to the moment of discovery. This method of sharing research does not necessarily preclude the referee process, which can be imposed through mediated online resources, but it does dispose of the delays and the costs of printing and physical distribution of the documents. In addition, the research will be seen by many more people than would be possible in a print medium, and because of the cost savings this medium can accommodate esoteric research that would never make it into print.

3. Too willing intermediaries

Libraries dispense knowledge and information they have amassed in their collections or, now, that they have otherwise gained access to through subscription or access fees. But by operating exclusively this way, as we do, librarians become extensions of the publishing industry. We get most of our information and knowledge resources not from the creators of the knowledge but from publishers, whether it be in print or electronic formats. Then we repackage and relabel it for our own customers.

This happens because for a very long time publishers were the most obvious source for knowledge and information. They were not the sole source for these commodities, nor the most economical perhaps. But because publishers had the distribution infrastructures, they were the most convenient source. They also could attract and hold the most notable writers, the researchers and the referees. Publishers, not the researchers or writers, and much less the librarians, controlled the release of new knowledge and information. Librarians simply made it available to wider audiences.

And we became very good at it, too, and very complacent. Perhaps the major intimidation of library automation has not been adjusting to the technology but to the false notion that electronic, non-print information resources were somehow upsetting the library's long-standing association with publishers. But that is not the case. We are still doing the same things we have done in the past — extending the reach of the publishing industry. Though the relationship between libraries and publishers has at times been strained, it became the convention and was therefore very comfortable.

4. Beyond format management

By using conventional print, CD-ROM and online resources, librarians are still middlepersons between publishers and our users. We still are too close to the publishers. But the possibility exists to be more direct intermediaries between our users and the real producers of knowledge. Using the new online technology, librarians can step closer to the sources of the knowledge and information they impart.

While continuing to manage various information formats, including paper, CD-ROM and online, libraries that are intent on promoting to the fullest extent possible a culture of research and learning among their users can exploit Internet/WWW technologies by creating their own local online databases that are specifically pertinent to their missions. Working directly with researchers, subject specialists and local information repositories, and by utilising Internet/WWW applications, librarians can create and manage specialised full-text resources composed of important current research and other information that otherwise would remain undiscovered by many persons in their user communities who need it. Once created on the World Wide Web and discovered by the growing number of Web-browsing researchers, these resources will attract online submissions written by other experts, faculty and specialists who wish to contribute to the growing knowledge base. These same contributors will be willing to regularly update their findings and engage in online forums and listservs, which will promote further research and the growth of knowledge.

Working with data communications specialists, libraries can make these databases accessible to widely dispersed researchers, students, practising professionals and the general public on local and international levels.

Online Information 96 Proceedings

Page 242

And by cooperating with businesses, professional associations and community organisations to whom they also make the local databases available, libraries can further enrich the content of these online resources and attract financial support for their continued development.

A number of library-based local online databases have already been launched. When I was last at the International Online Information Meeting, I spoke about the Human Genome Data Base (GDB) at the Welch Medical Library of Johns Hopkins University as an example of this new type of online resource (Ref 2). The GDB was designed with input from the library and is managed by librarians working with editorial boards composed of genetics experts. The database contains work from the ongoing studies of thousands of researchers and geneticists around the world (Ref 3). The GDB is not only a valuable resource but it has also set a new standard for research libraries. The GDB and other knowledge bases, as they are sometimes called, embody the methodology of the new librarianship, the *novum organum*, and represent the highest level of information, knowledge and research service to which modern libraries can aspire.

5. The new online

In just the last few years, however, the complexion of online information technology has changed completely, so that the new methods are available even to small libraries. No longer are giant mainframes necessary to host massive databases and search engines for researchers and other users who want the most current information in their fields. Super microcomputers, client/server technology and distributed computing configurations greatly simplify database construction, and make online resource development much more accessible and affordable for libraries. For instance, a high-quality, full-text WWW online database can be constructed for less than \$5000, including software that greatly simplifies Web page construction.

The new task facing a library that wants to increase its value as an information resource by producing and managing online databases is finding the 'hidden' information resources for the databases that would be appropriate to the library's mission. For this libraries must look outward to their communities, not to publishers. Academic and research libraries can begin within their own universities or research institutions among the faculty and other associated researchers and experts. After all, researchers, not publishers, are the source of the information libraries customarily add to their collections. And only a small portion of the research findings produced at universities and other academic institutions finds its way into print. The rest remains 'gray literature' that, while of great value to the right persons, never reaches the widest possible audiences.

Public libraries can look to schools, archival collections, local historians, ethnic associations, business and professional organisations, community actions groups, referral centres and other local information repositories to provide the content for valuable online resources. Every community provides unique opportunities to gather information, and the Internet/WWW technology provides an effective and affordable way to distribute it back into a library's community of users.

By exploiting Internet/WWW online technology and developing local databases, libraries can tap new fresh information sources and make them available to local audiences and others dispersed around the world in an online knowledge exchange. Libraries that do this will also induct librarianship into a new and extremely valuable endeavour.

6. The Kapiolani Library

The Library at Kapiolani College of the University of Hawaii has produced several full-text and image databases containing unique bodies of information that are pertinent to the mission of the Library and the College, and valuable to students, faculty and other users worldwide. Unlike earlier knowledge bases, the Kapiolani resources are Web-based, client/server databases. Some of them link the Library's Web servers with our CD-ROM server and online catalogue, and with related electronic resources across the world. Most of the databases were constructed with partner agencies that either were producing new information or research, or else had accumulated collections of knowledge or information that was not reaching the widest possible audiences.

Kapiolani's online databases are managed by the Library staff working together with our research partners in each of the separate projects. The new resources are bringing international recognition to the College, providing course enrichment for the curricula and serving a global audience of faculty, researchers, practitioners, students and the general public.

6.1. Asian studies curricular materials online

One year ago, the Kapiolani Library began a project with the Asian Studies Development Program (ASDP) of the East-West Center in Honolulu to convert a collection of curricular materials into a site for loading onto the Library's Web server. The East-West Center is a research agency of the US government that promotes cooperation and exchanges between America and countries in the Asian-Pacific region. The mission of the ASDP is to promote Asian studies in American higher education. Among its activities, the ASDP sponsors a number of annual Institutes to help faculty from colleges and universities across the US develop Asian studies courses and programmes. Institute participants create course syllabi and other curricular materials, drawing on what they have learned in the Institutes.

The ASDP staff had gathered these syllabi over five years in a growing collection of valuable Asian studies curricular models, and had transferred them to a disk collection, which was available to Institute participants and any other interested persons. The idea to make the collection into an Internet/WWW resource arose in a special workshop for librarians organised by Kapiolani Library staff for the 1995 ASDP Summer Institute in Honolulu, and the ASDP Web site was officially launched a few months later at a hardware and software cost of \$4700. The cost was covered by a research grant obtained by the Library to build the resource.

As a Web-based resource, the ASDP full-text syllabus collection reaches audiences well beyond the Institutes, and provides a wealth of models for Asian Studies courses in all fields. In addition, the database attracts submissions from Asianists around the world, which is helping the collection grow even faster. The ASDP staff performs the editorial duties related to the documents in the collection, and Kapiolani librarians maintain the hardware and software.

The ASDP online database, like our other databases, is keyword searchable, accepts online input from users and links to related Internet/WWW resources around the world. We use server software from the National Center for Supercomputing Applications (NCSA), and WAIS (Wide Area Information Servers) and Isite search engines for our databases. Working on these resources has not only upgraded the technical skills of the Library staff but has also helped us become aware of other little-known information collections in our area that could be converted into Web-based resources.

6.2. Emergency medical services data online

Kapiolani's Emergency Medical Services Database (EMSD) was also created within the last year to support the College's internationally recognised EMS Education Department. This Web resource provides direct access to the Library's CD-ROM server and medical CD-ROM indexes for document delivery services, and also hosts a growing collection of EMS research papers. This full-text collection is being developed to benefit EMS practitioners and students throughout Hawaii as well as the Asian-Pacific region.

For this full-text collection we are acquiring EMS information that is relevant to Asia and the Pacific from the US Department of Transportation, which has a vast collection of EMS curricular materials; from the EMS Clearinghouse in Florida, which also has a rich store of pertinent information; and from the US Federal Emergency Management Agency (FEMA), which also deals with EMS matters. We also solicit original submissions of research reports from EMS instructors, professionals and physicians in our hemisphere to create a truly unique database with a distinctly Asian-Pacific orientation.

The EMSD also supports specific EMS courses taught at Kapiolani College. One of them is a new graduate-level course to teach EMS administrators how to implement better emergency medical and disaster preparedness management systems. Partly because of the effectiveness of the EMSD, this course is being revised for distribution via the Internet and CD-ROM throughout Asia and the Pacific. The EMSD will be enriched to support this multi-national educational undertaking. There is a great demand for EMS information and know-how in the Pacific region. Developing nations there know that foreign tourists are more comfortable visiting countries that have excellent hospital and EMS systems in place.

Like the ASDP resource, the EMSD did not cost the Library any money. The Hawaii Department of Health contributed \$4400 for the purchase of the necessary equipment, including a high quality scanner. The Library created the Web site, and provides the expertise and staff time to manage it. The EMS Department at Kapiolani College performs the editorial work on the research papers, and provides general guidance about the type of content needed by EMS practitioners, which the librarians then find and add to the site. The EMS staff also maintains the home page's non-library information about the EMS programme, including course schedules, programme descriptions, admissions procedures, faculty lists, etc.

The EMSD project and the others as well have been marked by great enthusiasm in our partner agencies for the technology and what it can provide, and also for the Library for extending the technology to them. Even those who were sceptical of a project at the outset became ardent converts and firm supporters, much like librarians have been with automation in general. As a result, either the projects have come with funding or the partners have been willing to share the costs.

In fact, the EMS faculty were so enthusiastic about the EMSD that in addition to providing the Web-site development funds, they were instrumental in persuading the College administration to create a new librarian position specifically to manage and provide library services to the College's Health Education Departments. That was a battle I had been fighting single-handedly — and quite unsuccessfully — for years, until the EMSD project showed the faculty and the administration what a library could do with the right technology and the right personnel.

6.3. PRAISE

The Pacific Region Aquaculture Information Service for Education, or PRAISE, is a joint, federally-funded project of the Hamilton Research Library at the University of Hawaii's Manoa campus and the Kapiolani Library. PRAISE provides Internet access to a CD-ROM aquaculture index on the Kapiolani Library's CD-ROM server, document delivery of the items located through the CD-ROM index and a 'gray literature' database. The objective of PRAISE is to support the development of the aquaculture industry in the Pacific Region by promoting information transfer.

The Kapiolani Library developed and maintains the CD-ROM server and the PRAISE Web server that integrates these several services into a single Internet resource, and provides world-wide access to the site. The PRAISE document delivery tasks are performed by federally-funded staff at the Hamilton Library. The Kapiolani

Library was asked to participate in the project because of our sophisticated CD-ROM server and because of the Internet/WWW expertise resident on the Kapiolani Library staff. For its part in the project, the Kapiolani Library received \$3500 in federal grant funds for a new Web server, with additional funding likely in the future.

7. The Kapiolani Library Databases

The Web servers at Kapiolani Library have been online for a little over a year, and are accessed almost 15 000 times each month on average. Remote sites accessing our online resources include nations on every continent not covered by ice. PRAISE has the highest international access because it is the oldest of the resources with the most well-established user base. Next are the ASDP and EMS databases, which are newer.

The Library is also developing a number of online resources on the history and culture of Hawaii. These include collections of historical documents from the period of the Hawaiian Monarchy, as well as information on Hawaiian folk medicine and medicinal plants, foods of the ethnic groups in Hawaii, local artists and others. In these projects we are cooperating with museums, local ethnic groups, individual writers and researchers, archives and other libraries to develop and maintain the online resources.

These databases are unique in the world, and will provide important information services to our faculty and students while bringing international attention to Kapiolani College and Library. The ASDP database, EMSD and PRAISE were created with no equipment cost to the Library. Instead we received over \$12 000 in grants and other funding from our partners in the projects to develop the resources. This, in and of itself, indicates the ripeness of the information sources that are all around us, and the readiness of our users to have online access to this information, even though no conventional publishers would ever touch it. The Kapiolani local database projects also demonstrate the accessibility and affordability of the new online technology that can permanently revise librarianship for the 21st century.

8. The novum organum

This paper has presented some of Kapiolani's local online databases as potential models for the type of resource that libraries of the future can develop. Local online databases like these will greatly help libraries retain their crucial role in collecting and distributing current research in this era of electronic connectivity when the future of libraries is in question.

The most important advances in library science in the next decade and well beyond will involve the development of local online resources, which will soon become as distinguishing a feature of fine libraries as their print collections are now. These databases have greater flexibility of content than print resources and much wider accessibility. Overall, they are even proving to be more economical. Kapiolani's local online resources cost the Library nothing to create. In fact, they were created and developed in the midst of a severe economic downturn in Hawaii that reduced the Library's materials budget by more than 33%. The databases came into being in a year when we had no money to buy new books and were even forced to cancel over 100 subscriptions to periodicals.

Although we had long intended to create local online resources, I think it was partly the Library's inability to obtain new knowledge and information from our customary suppliers — conventional publishers — that motivated us to plunge forward into the new method of information gathering and distribution. The resulting accumulations of new knowledge and the active manner in which they are being accessed by our users have permanently changed the way we at the Kapiolani Library will perform our duties as librarians.

Terry D. Webb, PhD
Library Director
University of Hawaii
Kapiolani College
4303 Diamond Head Road
Honolulu
HI 96816
USA
Tel: +1 (808) 734 9267
Fax: +1 (808) 734 9453
E-mail: twebb@hawaii.edu

References

- [1] Jacobson, R. (1995) Researchers temper their ambitions for digital libraries, *Chronicle of Higher Education*, 42(13), A19.
- [2] Webb, T. (1992) Academic library super catalogues: reducing the research risks, *Online Information 92: Proceedings*, Learned Information, Oxford.



European Public Administration Information Systems on the Internet

María Jesús Fernández Ruíz

Centro de Documentación, Zaragoza, Spain

José María Angós Ullate

Centro Documentación Científica, Zaragoza, Spain

Isaac Bolea Aguarón

Random Centro de Informática, Zaragoza, Spain

Abstract: One of the principal challenges in public administration is adapting to needs derived from the application of new technologies that impact directly on managing as well as transmitting information and in supplying it accurately to the citizens. The Bangemann Report calls attention to the important role played in the development of the information society by public administrations who should dedicate serious efforts and resources to the innovation and development of information as an added value with democratising effects in society.

The current situation of European public administration information systems on the Internet is analysed through a statistical bibliographic study measuring the following parameters: producers; subject coverage; services, whether of added value or not; the legal framework; electronic publications; use of the classic documentation tools to structure information; use of GIS; database management; date of creation; update frequency; implementation of helps to locate information; confidentiality/privacy of the information; cost and quality control.

From the analysis carried out of European public administration information systems on the Internet, different conclusions are obtained that allow us to perform a more standardised design with improvements in the quality of the information and a better cost control. Analysis is performed of the use of systems which, based on those already existing and used in public administrations, permit the publication of this information on the Internet.

It seems to us that public administration should increase the use of new information systems such as the Internet. This will result in greater efficiency in the development of its functions and, therefore, in better service to the citizens, a service which constitutes the *raison d'être* of information systems for public service.

1. Introduction

Our present day society is the information society, where intelligence and knowledge are the resources of the future. This information society is built on telecommunication networks able to transmit great quantities of digital information. In June 1994 the Council of Ministers of the European Union, at the meeting in Corfu, considered the report of the High Level Group about Europe and the Global Information Society, known as the Bangemann Report, which calls attention to the importance of public administrations in the development of the information society.

One of the principal challenges facing public administration is to adapt to the new necessities derived from the application of new information technologies which directly involve managing as well as transmitting information, which must be supplied accurately to the citizens who ever more often ask for clear and concrete answers to their queries. These answers are not possible if we do not have true, updated and easily accessible information.

The relationships between administration and the administered are becoming those of server and client. Thus the Web servers of public entities and administrations are converted into authentic windows opening into cyberspace from which the public are supplied with information as to the services offered, or are even provided directly with those services, eliminating physical, spatial and temporary barriers in a way that would have been unimaginable only a few years ago.

The European Union has strongly supported the development of the field of telematics with programmes such as IMPACT, which ended in December 1995, and the INFO2000 programme currently operating.

It is clear that public administrations cannot be excluded from this process and in fact they have not been, as

can be seen in the high number of servers with all sorts of information that have been generated.

The motivation for this inclusion of information in the Net comes from different objectives:

- to provide access to the important information resources available, statistical as well as bibliographic and documentary;
- to give information about the departmental and organisational structure of central, autonomous and local administration;
- to divulge information about the programmes and initiatives that the different administrations may establish or bring together;
- to facilitate communication with the citizens as well as with other public organisations, with a view to greater administrative simplification.

One of the ways that public administration can improve service to the public is to provide the information. An information system on the Internet permits public administration:

- to eliminate physical distance between the seeker of information and its source, that information being accessible through any personal computer anywhere;
- to facilitate all types of external information to the administration, instantly and without needing to telephone or write to other national or international centres;
- to offer the user the availability of information 24 hours a day.

2. Directory of European public administration information systems on the Web

By European public administration we mean:

- (a) the general administration of each state;
- (b) the administrations of the autonomous communities;
- (c) the institutions that form the local administration.

The construction of a directory of European public administration information systems on the Web is in an initial phase of development. Its principal purpose is to make known the existing resources in this matter, to promote their use and to analyse the problems found by administrations in putting their information on the Net. In spite of being in the development stage, we have reached the conclusion that these are today very heterogeneous.

To collect the data that compose this directory, the Lotus Notes database management system was used. The information collected for each information system permits the description of their most important characteristics: producers; subject coverage; services, whether they are added value or not; legal framework; electronic publications; utilisation of the classic documentation tools to structure the information; utilisation of GIS; database management; date of creation; updating frequency; implementation of support to locate information; confidentiality/privacy of the information; and cost/quality control.

One of the tools we used for the statistical bibliography study of European public administration information systems on the Web is a NETCARTA spider called CYBERPILOT (TM) PRO. Before the use of the spider the addresses of the Web sites were located using different search robots: Alta Vista, InfoSeek, Lycos, Ole, Ozu, Webcrawler, Yahoo and so on. The addresses provided must then be analysed and information is obtained from each Web site that allows us to conduct the study.

We are aware that there are mistakes or omissions in spite of the efforts made to complete the information: we expect them to be corrected during the development of the project.

3. Current situation of European public administration information systems on the Web

The problems facing public administrations are on the one hand the safety, confidentiality and privacy of the information and, on the other, the shortage of both hardware and software infrastructures which is the reality in public administrations and which makes it difficult for them to export information to the Net.

The development of applications by public administration has been based on powerful database environments that permit the management of large quantities of information but which are today very complicated to export to the Web. Rewriting the information to produce HTML (hypertext markup language) pages is too tedious a task to be considered and, furthermore, all the functions which a search motor offers are lost.

Moreover, it is the intention of public administrations to use Internet applications within their own Nets to create an intranet that permits them to distribute information internally at a cost which is not excessively high.

Below some solutions are described which have been adopted by public administration to create a Web site on the Internet, and other more recent ones that we believe can be recommended and that will be used in the near future. The trend on the part of the suppliers of databases was to deliver a series of CGIs (Common Gateway

Interfaces) that permit the construction of a gateway between the relational database and the Web server. These CGI's can be classified in two categories: the developer of the CGI inserts SQL statements in it to work with the databases; and some products possess CGI standards which, installed in the Web server, constitute the interface with the DBMS (database management system).

Among the databases that provide this sort of software we may mention Informix, Oracle and BRS.

At the same time, Sun Microsystems with its Java language is working toward a standard similar to Microsoft's ODBS to work with the databases. This supposes that an application written in Java will be able to exchange data with thousands of applications already existing, and it will be able to generate what has been called the data Web in which the browser is only a front-end between the database and the client. Currently we can develop an interface between the program written in Java and the API of the database, but Sun is developing the JDBC (Java database connectivity).

The JDBC is a middleware level that presents a standard API for access to the data (to open a connection, to launch a conference and so on) which will load the necessary driver for the specific database. Some suppliers of databases have already announced support of the JDBC protocol: Oracle, Gupta, Informix, IBM with DB/2, Object Design and Sybase.

They will provide Java extensions to their tools but the real step forward would be to propose tools completely developed in Java. In addition to this, a final phase would be to give to Java surfers the possibilities provided by transactional monitors, permitting evolved mechanisms of data validation, managing a considerable number of transactions and the possibility of access to distributed databases.

On another front we find some suppliers of databases such as Oracle and Lotus which, through their products Oracle Web Server and InterNotes Web Publisher, provide all the power of a database server integrated with a Web server.

Oracle Web Server is an integrated and complete environment with Oracle 7 as the engine of the central database. With this system applications can be created in PL/SQL (the procedural extension of SQL). The hypertext links contained in any HTML document can now directly run a stored procedure, allowing Oracle 7 to create dynamic HTML documents. The software packages that compose Oracle Web Server (Oracle Web Listener, Oracle Web Agent and Oracle 7) are totally configurable from an HTML interface. This solution of Oracle is totally compatible with HTML 2 and the improvements proposed by HTML 3.

As far as Lotus is concerned, it presents a solution to integrating intranet with Internet. Basically the solution has two parts: firstly there is Lotus Notes 4.0, database and groupware application that now includes a gateway with e-mail, SMTP and X400 as well as a database called InterNotes Web Navigator that permits the users of Notes to navigate the Web with their own Notes interface. This software also permits access restriction to certain Web sites, and to store and update the sites visited most frequently by our clients in order to be able to access the Internet with more agility.

Secondly, Lotus has a whole range of products known generically as InterNotes that permit the automatic publication of all our database Notes on the Internet. The basic operation of InterNotes is to create Web pages from a Notes database and to translate their documents to HTML, including the images that may be found within the documents. The update of these pages is performed automatically and periodically, at time lapses which may be as short as 30 seconds.

To carry out searches InterNotes uses the Notes server, thus being able to seek with all the complete text search functions offered by Notes. Access to the documents is through the views in the Notes database (indices ordered and categorised by entry fields to the documents) creating HTML documents of such views and updating each time the database is updated. Furthermore InterNotes Web Publisher provides Java and VRML support.

At the same time Lotus is working on new Web server software called Domino that will permit direct access to Notes servers and other database servers.

4. A guide to the use of the Internet in public administration

Public administrations now need to learn how to implement their initiatives on the Internet and how to distribute their information with security, and also need their documents to have a common basic format. It is for this reason that certain minimal procedures will have to be established which will be followed by all administrations:

- (1) The Internet is a complex phenomenon that brings together different disciplines such as telecommunications, public relations, high technology, information management, security and so on. Its use by public administrations will mean incorporating these disciplines in the design of an information system if it is to be successful. Therefore its design will have to consider the creation of a multidisciplinary team with representatives from different departments: documentalists, telecommunications, the press and administration.

The Internet is only one of many mechanisms that a department can use to communicate with the citizen. Departments should consider setting up a process to identify and standardise the material being distributed on the Internet in the same way as they do with other publications.

- (2) The costs of using the Internet include the time and resources needed to create and maintain information, the resources required to train personnel, telephone connection charges, and hardware and software costs. How can departments justify these costs? Is the Internet actually necessary today or should we work with it, waiting for it to be essential in the near future?. The implementation of the

Internet in all public administration departments requires high investment. To verify the costs and benefits of its use, a pilot experience can be carried out in one department which serves as a model to the others; limited access to e-mail or FTP could be offered to its users; or the number of users be limited. The departments would have to access the Internet through local nodes to avoid long-distance charges. The time and resources required to transform the documentation available in the Net into digital format should also be analysed.

- (3) Equality of access must be guaranteed. Until access to the Internet is ubiquitous, public administration must still be made available in other formats for those who do not use or do not have access to the Internet. Information on the Internet exists in digital format, thus giving visually and hearing-impaired clients better access to information than many traditional formats and hence satisfying the right to information of all citizens.
- (4) Integrity of the information transferred from the administration to the citizen. This can easily be changed and reposted to anyone else without the administration's knowledge or consent. Electronic information is easy to copy and modify. It is very difficult to avoid this and, therefore, it is important that those posting the information be aware of the time and date of the last update posted, and introduce a contact name and number so that users can verify the accuracy of information. This is particularly important for information relating to health and safety.
- (5) Because of the nature of the Internet (a cooperative network with each node relying on the next for service) there is currently no service guarantee. Moreover, once a particular packet of information is in transit from a department to a user, we cannot predict the route it will take and the time it will take to arrive. Departments should determine the level of priority of service required and have alternatives in order to guarantee the result.
- (6) Internet e-mail is not private. At the present time there is no foolproof way to authenticate the sender of a message or to be assured that an e-mail message was not intercepted and changed. Communications via e-mail should reflect the same tone and manner used in traditional correspondence. The e-mail policy of departments should indicate that classified and extremely sensitive information must not be transmitted via e-mail unless encrypted or protected properly.
- (7) Personal information exists electronically within the public administration information systems connected to the Internet. From a security perspective, internal information systems containing personal information should be separated by a firewall (a specifically programmed machine that restricts access by controlling the interface between a subnetwork and Internet) or be located in a physically separate server that the public cannot access via the Internet.
- (8) Users of public administration information on the Internet need to know that they can locate the information they want and that the information they locate is up to date, accurate and authentic. For this reason of identification, a similar presentation and format of all documentation in the Web is necessary. If departments provide a uniform description, consisting of a few basic elements, with each document on the Internet, users will be able to identify the relevant information in the documents rapidly and easily. For example, each Web page must include the title, creator, date, registration number, availability, access, publisher and URL, among other details.
- (9) Public administration departments should consider security of information and departmental resources before connecting to the Internet. Security safeguards will differ depending on whether the department is providing internal access to the Internet, including e-mail, or external access by the public to departmental information, or both: the use of passwords, firewalls or gateways that isolate the departmental systems using computers which connect one network with another when the two networks use different protocols.
- (10) The elements should include a Web page in the information system of the public administration in order to provide the Web surfer with correct identification:
 - *Title*: this is the name that distinguishes the documents. It appears at the head of the document;
 - *Creator*: identifies the author of the document. The name of the department, the administration and the country should be included. It can include the place where it is filed and the individual author's names and even the unit or service within the department;
 - *Date*: in day/month/year format. It is the date on which the document was published. The last review date should be included so that users can determine if it has been modified;
 - *Number of record*: ISBN, ISSN, URN if it is known;
 - *Validity*: describes any legal limitation, if one exists, on the use or distribution of the document;
 - *Access*: size of the document, existence of tables, images or sounds, hardware or software requirements to access the document;
 - *Publisher*: name of the supplier of the document;
 - *URL*: Uniform Resource Locator (URL). This provides the name of the host, port, path and filename needed to retrieve the document;

- **Abstract:** summary of the content of the document;
- **Key words:** descriptive terms that summarise the content of the document. They would have to be assigned in the different official languages of the European Union.

5. Conclusions

We have verified that Internet implementation is not homogeneous in all countries and public administrations.

The public sector is beginning to use the advantages of the Net to broadcast its messages and to offer citizens useful information about the administration. Public administration has extensive resources of information that are very valuable for individual users as well as for the community. It can supply information to the business world as well as to the suppliers of information who give services of added value, and to society in general. However, access to this information tends to be difficult and cumbersome in all the member states of the European Union. It is because of this that better access and use of the information resources of the public sector should be provided. We believe that the creation of the directory proposed in this paper will help the citizen.

We are of the opinion that public administration should increase the use of the new information systems such as the Internet since this will result in greater efficiency in the development of their functions and, therefore, in a better service to the public — a service that constitutes the goal of information systems for public service.

As we have previously said, what is necessary is the existence of a methodology — standards on the part of the European Union — to create a public administrations network on the Internet, and it would be necessary to create programmes for regional development because the situation at the moment is not homogeneous.

We know that the existence of standardisation permits the availability of rapid, reliable information, and the improvement of relations between the administration and the citizen and of the services that supply the administration. We are aware that a general plan of action must be developed that guarantees quality.

José María Angós Ullate
 Universidad de Zaragoza
 Centro Documentación Científica
 Edificio Geológicas
 Ciudad Universitaria
 50009 Zaragoza
 Spain

References

- Bangemann, M. (1994) Europa y la Sociedad global de la información. Recomendaciones al Consejo Europeo (Recommendations to the European Council Europe and the global information Society), *Brussels Meeting of the European Union. Corfu, 24-25th June 1994*, Brussels, 26 May 1994 (CD-84-94-290-ES-C).
- Bangemann, M. (1995) *Libro Verde de la innovación*, COM(95)688 final, vol. I, II. Brussels, 20 December 1995.
- Bellovin, S. (1989) Security problems in the TCP/IP protocol suite, *Computer Communication Review*, **19**(2), 32.
- Chaléat, P. (1996) *HTML y la programación de servidores Web*, Editions Eyrolles, Paris.
- Cerf, V.G. (1993) On national information infrastructure, *Bulletin of the American Society for Information*, **20**(2), 24.
- Cohen, P.A. (1996) *Lotus Ships InterNotes Web Publisher Release 2.0*, Lotus Development Corp., Cambridge, MA.
- Curry, D. (1992) *Unix System Security*, Addison Wesley Publishing Co., Inc., Reading, MA.
- Gallardo, M. (1996) *Seguridad en Unix*, Editorial Paraninfo, Madrid.
- Huitema, C. (1995) *Internet... una vía al futuro*, Editions Eyrolles, Paris.
- Macaray, J. (1996) *Programación Java*, Gestio 2000 Edicions SA, Barcelona.
- Morant, J.L. (1994) *Seguridad y protección de la información*, Centro de Estudios Ramón Areces Editorial SA, Madrid.
- Negroponte, N. (1995) *El Mundo Digital*, Ediciones Zeta, Barcelona.
- Otte, P. (1995) *La Superautopista de la Información, más allá de Internet*, Prentice Hall, Englewood Cliffs, NJ.
- Ranum, M. (1994) Internet firewall protection, *Open Computing*, **11**(9), 95.
- Rodríguez de las Heras, A. (1991) *Navegar por la información*, Fundesco, Madrid.
- Scotkin, J. (1996) Negocios con Java, *Datamation*, **122**, 30-31.
- Sterne, J. (1995) *World Wide Web Marketing, Integrating the Internet into Your Marketing Strategy*, John Wiley & Sons, New York.
- Sun Microsystems Ibérica (1996) *Internet y las empresas: Hacer negocios en la red*, Grupo Negocios, Barcelona.
- Terceiro, J.B. (1995) *Sociedad Digital: del homo sapiens al homo digitalis*, Alianza Editorial SA, Madrid.
- Varney, S. (1996) Enlazar el Web con datos y aplicaciones heredados, *Datamation*, **123**, 82-92.
- Varhol, P. (1996) ¿Quién gana el duelo entre servidores Web? *Datamation*, **123**, 94-99.



The delivery of business information to the end-user — user perceptions and needs

Alma Swan

Perspective Business Associates, UK

Sheridan Brown

Key Consulting, UK

Abstract: End-users are a rapidly growing segment of the market for academic and commercial information. Increasingly, end-users are making buying decisions — choosing the information they want, selecting delivery options and deciding on preferred payment methods. The key success factor for all parties in this value chain, from information producers and publishers to third parties delivering this information to the customer, is understanding end-user perceptions and needs with respect to sourcing and using work-related information.

Our studies of two groups of business information end-users — in the business education and commercial sectors — have revealed interesting insights into the perceptions of these potential customers. Intriguingly, there are often marked differences in attitude between the groups, in terms of both desire for information and willingness to put effort into accessing it. These and the other key issues we have identified — attitudes to electronic delivery, ease of access versus content and perceptions regarding future requirements — have important implications for information providers undertaking product development and sales and marketing planning for this sector.

Keywords: end-user, business information, electronic publishing, Internet, market research, market segmentation, product development

1. Introduction

When Day *et al.* spelt out their customer-oriented concept of what constitutes a competitive market, they argued that people seek the benefits that products provide rather than the products *per se*, and that it is the usage requirement which dictates the benefits being sought (Ref 1). While most modern marketers would concur with this view that customer needs and requirements should be paramount when developing new products, great care is still required not to slip into the mode of technological determinism where new products are created simply because new technological advances enable their development. The need to understand the requirements of targeted users remains a very important aspect of the new product development process and is therefore an important objective of the research reported here.

We elected to survey customer needs and information-usage behaviour in two sectors, both using business information. The corporate sector's consumption of, and attitudes to, information has long been the subject of examination and the overall concept of information as an asset is familiar at company level (Refs 2, 3). In this paper we report the attitudes and understandings specifically of end-users in the corporate sector. The growth of electronic access and retrieval of information may be influencing the information-seeking and usage behaviour of these people, which in turn will affect aspects of information purchase by corporations.

We have also probed the attitudes of academic end-users who are utilising business information. Information-sourcing has always been part of academics' professional skills and traditional library work has an integrated role in the academic way of life. For this group too, though, new access and retrieval methods are changing the way things are done and other influences, notably the high price of serials and consequent reduction in library subscriptions, are giving rise to new buying patterns and demand for different product features. Whilst several recent studies have focused on academic end-users in the sciences, the area of business education may have its own attributes and characteristics (Refs 4–7). This study investigates this, and reports on these business end-user market segments and the implications for information providers.

2. Methodology

- (a) *Corporate end-users.* A questionnaire was distributed to 198 named senior managers of a large UK company in the energy sector. The questionnaire was accompanied by a covering letter from the appropriate Director, expressing support for the investigation and encouraging the managers to respond. Responses were returned to the Director's office and passed to us for evaluation. Of 198 questionnaires distributed, 97 were completed and returned (49%).
- (b) *Academic end-users.* Questionnaires were sent to 595 faculty members of 11 top business schools. Nine countries were represented in the survey (Australia, Canada, Denmark, Israel, Japan, the Netherlands, South Africa, UK, USA). Of the 595, 445 were named individuals: a covering letter explaining the nature of the research was included. Responses were received from 59 people (10%).

The questionnaires were designed to probe the extent of end-user access to, and usage of, work-related information and the perceived value of such information in facilitating their work. The questions were slightly different on the two types of questionnaire, reflecting the differences in information provision and access in the two types of workplace. Where questions contained different wording but explored comparable situations, it is made clear in the results and discussion sections below.

3. Results

3.1. Profiling

The questionnaires asked about the seniority and length of service of the respondents, and it was clear that within the corporate respondent group two discrete subgroups could be identified. The *high flyers* group comprises managers who are below the age of 35 and have already achieved relatively senior roles in the company. The implicit assumption is that reaching such high ranks at a relatively young age marks out these individuals as achievers. Eight managers fell into this category (8% of the total population). The *stable core* group comprises managers who are 46–55 years old and have reached middle to upper levels of senior management. The assumption is that, while competent, these managers are unlikely to achieve the highest management grades during the remainder of their career. Seventeen managers match this profile (18% of the total population). Where these two groups differ in response from the full population it will be a matter for discussion below. Of the academic end-users, 20 (34%) had less than 10 years service and the other 66% represented senior faculty members. There were no significant differences in responses between these two groups so for the purpose of this study all academic respondents are treated as one profile group.

3.2. Information access and usage

The first major issue was whether end-users in the two work environments have access to the information they want. Despite the fact that all or most of the journals required by academic respondents are available in their library (76% agreed with this statement) and access to the peripheral sources they value is easy, 58% still feel they are missing useful information and use other available sources to try to find it. Corporate end-users are in agreement, with 86% agreeing that better access to external information would improve the performance of their department. The high flyers show a particularly strong agreement with this statement. This response should be appraised with caution, however, since there seems to be a discrepancy between access and use, at least in some cases. For example, where managers have access to information on CD-ROM, less than half actually use it. Conventional online- and Internet-delivered information follows the same pattern. In the academic situation, too, a huge proportion of respondents have access both to Internet facilities (in addition to e-mail, which is used by all respondents) and to online services provided via their library, yet usage figures for these media are quite low. Our corporate respondents indicated that they are not confident of their abilities to use these media effectively and anecdotal evidence suggests that academic end-users feel similarly.

We also asked respondents to rank the sources of information according to importance. Unsurprisingly, academic end-users rated journals as the most important source of information (86%): undertaking a systematic search of these journals, in printed form, in their institutional library was the main method of access. Thirty six percent of respondents in this category also considered browsing through printed journals in their library to be extremely or very important to them. Fifteen percent report accessing primary journal information electronically in their library. Local access (in their office) is also high, with 49% having personal subscriptions to at least one journal and 44% having access to (and using) departmental copies. Secondary information (abstracting and indexing services) is rated highly in importance, with 51% using these services in print and an even higher percentage (58%) using them on CD-ROM in their libraries. CD-ROM is also a popular medium in the local environment for secondary information (31%); the usage of electronic media is discussed separately below. Secondary information in print is used locally by 34% of respondents (Table 1).

Table 1: Types of information rated extremely or very important for their work by academic respondents. The figures refer to percentages of total respondents.

Information usage by academic respondents	
Information type	%
Primary journals: systematic library search	86
library browsing	31
personal copies	49
department copies	44
colleagues' copies	24
online from own desk	15
Secondary info: library print copies	51
library CD-ROMs	58
library diskettes	15
library online	20
print at own desk	34
CD at own desk	31
diskette at own desk	14
online host from own desk	12
online to library service	51
Inter-library loans	75
Authors' reprints	14
Other libraries	32
General informal	76
Conferences and meetings	41
Other sources: newspapers, magazines (usage)	76
broadscope journals (usage)	66
trade magazines (usage)	54
TV/radio (usage)	53
Internet newsgroups (usage)	31
rated as highly important by	58

Whilst more peripheral sources (newspapers, trade magazines and broad-scope publications) are ranked extremely or very important by a large proportion (58%) of academic end-users, they are given top ranking by 94% of corporate personnel. This group also ranked what we termed word of mouth information as their second most important category of information (see below), with journals in third place (Table 2). When asked which they considered to be most accurate and reliable, however, journals win the overall vote, though the high flyers subgroup rate consultancy most highly.

Table 2: Information usage by corporate respondents. The first column of figures shows the responses to the question 'In terms of the sort of information you require for the normal performance of your duties, which of the following external information sources do you have access to?' Figures refer to the percentage of total respondents. The right-hand column shows these sources, ranked according to importance, when respondents were asked which they actually used.

Information type	Access to information type (%)	Ranking (% giving this rank)
Newspapers and magazines	96	1 (94)
Journals	92	2 (91)
General informal	87	3 (84)
Conferences and meetings	90	4 (84)
TV/radio	83	5 (65)
Consultancy	58	6 (58)
Company reports	65	7 (52)
Market reports	56	8 (48)
Abstracts	53	9 (44)
Online	37	10 (25)
Electronic bulletin boards	37	11 (23)
E-mail	15	12 (10)
CD-ROM	11	13 (5)
Internet	4	13 (5)
Microfiche	4	14 (2)

3.3. Informal or oral communication

Oral or informal written communication (we termed this 'word of mouth' in our questionnaire) does seem to have a high weighting as an important source of work-related information. Under this category, the *corporate high flyers* rate conferences the most useful type of information source while the stable core group selected general informal communication channels. The high flyers also gave a very high rating for accuracy and reliability to consultancy. Seventy six percent of academic end-users also say that general informal communication is extremely or very important, and 41% of them rate conferences, meetings, seminars and other informal communications with non-local colleagues in the same way.

3.4. Electronic information

The questions asked about in this section were designed specifically to examine the attitudes of managers and academics towards the use of electronic information products, and whether they have the skills necessary to use such products.

One hundred percent of academic end-users have an e-mail facility, 88% have access to other Internet services as well and 80% have access to online information sources via links to their library. A small proportion (15%) actually use the latter to access primary information and 12% access secondary information (abstracting and indexing services) this way. Twenty percent access online services from their library itself (this may be with help from an information professional). The biggest category of electronic information usage was for secondary information provided on CD-ROM in their library (58%) closely followed by accessing this information via a link from their desk to their library: 51% of respondents use such a service (Table 3).

Table 3: Internet usage by academic respondents. Figures give percentages of respondents in each category.

Access to and usage of the Internet (% academic respondents)	
Access available	88
Primary journals accessed via the Internet	20
Secondary information accessed via the Internet	19
Would like information delivered via the Internet	98
Would expect institution to pay for information delivered via the Internet	98

Electronic information usage in the corporate sector differs in detail but is equally diverse. Access to online databases, the Internet and CD-ROM products is seen as a major source of competitive advantage by 66% of the full population of respondents, though the high flyers tended to agree rather than strongly agree with this statement. A follow-up question asked whether respondents felt they could use these electronic resources effectively: the answers showed conclusively that managers do not feel confident in their abilities in this area. The 'no' category was ranked first in every case, with a few respondents opting for the 'somewhat' category (Table 4). In the case of online databases, technology that has been available to this sector for a number of years, confidence was generally higher than for CD-ROM and the Internet. A fair proportion (38% of the high flyers and 27% of the full population) selected the 'somewhat' category for this medium.

Table 4: Corporate respondents' perceptions of their ability to use electronic information sources. Figures give percentages of total respondents.

Ability to use electronic information sources (% corporate respondents)			
	Yes	Somewhat	No
CD-ROM		16	71
Online		27	49
Internet		20	67

Finally, and in contrast to the above findings in the corporate sector, the academic respondents were overwhelmingly in favour of greater access to electronic information. Ninety eight percent of them would like work-related information provided via the Internet: the same proportion expected their institution (library or department) to pay for information delivered by this channel. This is interesting in the light of our findings regarding personal subscriptions to printed information sources given above (see Discussion below).

3.5. Information purchase

We asked the academic respondents to indicate how many personal subscriptions they had to journals in print and the results are shown in Table 5. Forty six of these are paid for by the individual concerned and 36% are purchased using departmental funds. Thirty six percent of respondents have been put off taking out a personal subscription by difficulties in the payment process.

Table 5: Information purchase patterns by academic end-users. Figures refer to percentage of total respondents in each category.

Information purchase by academic respondents	
Number of personal subscriptions:	
0	19
1	15
2	15
3	17
4	7
>4	27
Payment method:	
self	46
department	36
grant	7
share with colleagues	2
Respondents who have been put off a personal purchase by subscription difficulties	36

4. Discussion

In both academic and commercial environments, information which is relevant, accurate and timely is one of an organisation's chief assets. In both situations, too, the main drivers of change are the price of information and technological developments affecting available methods of delivery. Publishers seeking to maximise opportunities in the business information market have two main strategic pathways open to them. The first is to develop new markets, and the most attractive option in this respect is the end-user sector which has been largely untapped except by publishers of broad-scope and consumer-type information. The second is to develop new products: under this heading electronically published information, which leverages traditional publishing strengths with the power of new technology, appears to be an avenue to business success.

We undertook this survey in order to probe more deeply into end-user perceptions and attitudes with regard to accessing and using information for their work, with the aim of drawing conclusions which pertain to the way publishers may develop their business in these markets in the future.

The academic respondents acknowledge that their libraries provide access to the majority of learned journals they require, but more than half the respondents still feel they are missing information which would be useful. Broad-scope and other peripheral information sources are searched out and channels such as inter-library loan, other libraries or authors' reprints are commonly used. This desire for more access to useful information was echoed by our corporate respondents, who were emphatic that better access to information would improve performance.

Anecdotal comments on the returned questionnaires indicate some level of anxiety on this topic, with respondents declaring their willingness to employ all channels to get to useful information; their main problem is awareness of its existence in the first place. Clearly, they perceive an important gap here between information available somewhere and their awareness of that fact. Additionally, the issue of 'information overload' does not seem to pertain in the situations examined here: both managers and academics appear eager to access and use as much information as possible (Refs 8, 9).

When questioned specifically about their attitude to information delivered electronically, both groups responded positively. The corporate respondents agreed that greater access to these sources would be a major source of competitive advantage while academics overwhelmingly agreed that to be able to access the information they needed via the Internet would be advantageous to their work in the future. These responses are made emphatically despite the fact that these services are already available in many cases. Although there is a palpable growth in interest in using the Internet in corporate situations (Ref 10), a large proportion of our corporate respondents indicated they currently do not use this medium to any extent. Since end-users are keen to take advantage

of new electronic delivery methods but are not actually employing them when they are available, we suggest the existence of a training gap in this area. This is borne out by our findings that the corporate respondents admit they cannot use electronic media effectively. Publishers may capitalise on this and drive the market forward by providing simple-to-use electronic products which overcome the difficulties and incompatibilities of present offerings.

Another issue with respect to electronic delivery concerns the use of what we termed broad-scope and peripheral information. This category is very important to both academic and corporate end-users, with both groups consulting newspapers and magazines for work-related information. Whilst at present these are still almost all used in print, the availability of such publications on the World Wide Web is growing fast, with most of the world's major broadsheet newspapers and a significant proportion of specialist and trade magazines fast establishing a presence. The opportunity for librarians to add value to these services is also being exploited (Ref 11). Increasing familiarity with these electronic products, provided in the Web's open environment, will serve to push end-users rapidly up the learning curve in electronic access and retrieval, and should provide the opportunity for publishers of other types of business information to capitalise on the advantages of Web publishing.

The importance of informal communication should not be surprising in the light of previous findings on this topic. A Royal Society survey of information usage in science, technology and medicine in the UK found that personal contact was used for information by 97% of respondents, and two recent studies of biological researchers found that discussion with colleagues or conferences was used to gather information by up to 95% of respondents (Refs 4, 9, 12). The high rating for consultancy by the corporate high flyers took us slightly by surprise, but presumably in a work environment with severe time pressures the traditional role of consultants in distilling, synthesising, researching and interpreting information for subsequent managerial use becomes invaluable.

Finally we raise the issue of purchase decision-making. In the academic environment it is clear that while major spending decisions are still a matter for library budgets, individuals are taking advantage of special rates for print subscriptions and there would seem to be little reason to suppose that a similar take-up would not be found for electronically-delivered information if priced correctly. Publishers should note, however, that a sizeable proportion of respondents declared that they had been put off a personal subscription purchase by insurmountable difficulties in the payment process (we asked if they had actually been put off the purchase rather than just whether they had experienced difficulties). Efficient and straightforward subscribing procedures must be in place if publishers are to capitalise on end-users' readiness to pay themselves for work-related information.

The Internet rather confounds this issue. In answer to our questions about electronic access via the Internet, all respondents who indicated that they would like to receive their information this way also stated that they would expect this information to be paid for by their institution. It seems that Internet-delivered information is perceived as 'institutionally-provided', on a par with information delivered to libraries on tape or CD-ROM and networked campus-wide, and probably reinforced by the runaway success of services such as BIDS (Bath Information and Data Services). A recent survey of academic end-users in biomedicine by one of us (APS) found that such services, where information is free-at-the-point-of-use but purchased via one decision by the librarian, constitute an increasingly popular category of information accessed by end-users (Ref 12). The growth of such services, especially when provided by intermediaries, is clearly of fundamental importance to publishers, encompassing as it does such issues as the pricing of networked electronic products and the scope for publisher-developed product features to give a sustainable competitive advantage.

In conclusion, we see servicing the end-user market as a huge opportunity for publishers and one which is currently hardly exploited at all in the sectors we studied. Clearly, though, just as there are many differences between the needs of end-users and those of institutional buyers, so there also appears to be significant and often substantial differences in the perceived needs of different types of customers in the end-user sector. Careful segmentation of this market will be necessary if it is to be served efficiently and profitably.

Alma Swan
Perspective Business Associates
The Leys
Gaddesby
Leicestershire LE7 4XF
UK

References

- [1] Day, G.S., A.D. Shocker and R.K. Srivastava (1979) Customer-oriented approaches to identifying product-markets, *Journal of Marketing*, **43**, 8-19.
- [2] Waddington, P. (1995) Information as an asset: the invisible goldmine, *Business Information Review*, **12**(1), 26-36.
- [3] Porter, M.E. and V.E. Millar (1985) How information gives you a competitive advantage, *Harvard Business Review*, July-August, 149-160.
- [4] Bukhari, A.A. and A.J. Meadows (1992) The use of information technology by scientists in British and Saudi Arabian universities: a comparative study, *Journal of Information Science*, **18**(5), 409-415.

- [5] The Royal Society, The British Library and The Association of Learned and Professional Society Publishers (1993) The scientific, technical and medical information system in the UK, *British Library R & D Report No 6123*, London.
- [6] Meadows, A.J. and P. Buckle (1992) Changing communication activities in the British scientific community, *Journal of Documentation*, **48**, 276-290.
- [7] Rolinson, J., H. Al-Shanbari and A.J. Meadows (1996) information usage by biological researchers, *Journal of Information Science*, **22**(1), 47-53.
- [8] Levitt, T. (1987) From the Editor, *Harvard Business Review*, January-February, 4-5.
- [9] Katzer, J. and P. Fletcher (1992) The information environment of managers, *Annual Review of Information Science and Technology*, **27**, 227-264.
- [10] Foster, A. (1995) Using the Internet for business information, *Business Information Review*, **12**, 2-17.
- [11] Smith, G. (1995) Access to newspaper collections and content in a time of change, *IFLA Journal*, **21**(4), 282-286.
- [12] Swan, A P (1996) Information usage by biomedical scientists (unpublished observations).



Towards real information on demand

Philip Barker

Human-Computer Interaction Laboratory, University of Teesside, UK

Abstract: The phrase 'information on demand' is often used to describe situations in which digital electronic information can be delivered to particular points of need at times and in ways that are determined by the specific requirements of individual consumers or client groups. The advent of 'mobile' computing equipment now makes the realisation of this technique a practical reality. In order to discuss this approach to information provision, this paper classifies digital information archives into four basic types: global, local, isolated and mobile. The characteristics and properties of these different types of information are described and the relationships between them are then discussed. The taxonomy that is introduced is used as a basis for the presentation of a case study which illustrates some of the ways in which this type of equipment can be used to access, update and process remote online information in a totally interactive fashion.

Keywords: decision support, digital data archives, flexible access to information, information on demand, mobile computing, notebook computers, online information, personal digital assistants, problem-solving

1. Introduction

As a consequence of the various evolutionary changes that have taken place, the human species has risen to a position of uniqueness and prominence within the hierarchy of biological systems. The important developmental changes that have occurred during its evolution have resulted in the creation of a sophisticated psychocognitive system of control. Because of the complex nature of this control system, virtually all cognate human activity is governed by innate goal seeking behaviour. For example, the need to work, eat, communicate, learn and relax are all 'driven' by inherent goals that are derived from our sophisticated physical and cognitive make up.

Naturally, the realisation of any particular goal (within any given situational context) will usually involve the successful solution of an intricate pattern of problem-solving activities. The exact nature of the problems to be solved in any given situation will depend critically upon a number of generic control factors. Some of the more important of these are: the required outcomes; the quality of the solution necessary; resource issues (for example, finance, person-power, equipment and information availability); time scales; the problem setting; and so on. Fundamental to the appropriate use of resources to solve a problem is the associated decision-making activity that takes place. Ideally, this should lead to a solution that falls within the 'region of acceptability' for the problem being solved — bearing in mind the constraints imposed by the available resources, the time scales involved and the assumptions that are made ('the givens') at the onset of problem-solving activity.

Obviously, successful (and optimal) problem-solving depends critically on the quality of the decision-making that is used to pick out the most suitable pathway through a problem space. That is, the path from the initial state to the final state of problem-solving. Three different routes through a simplified problem space are illustrated schematically in Figure 1.

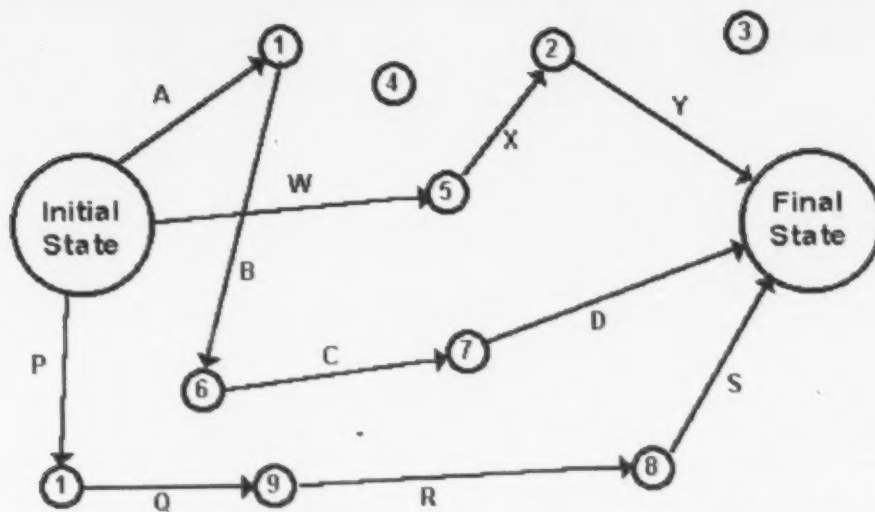


Figure 1: A problem-solving graph.

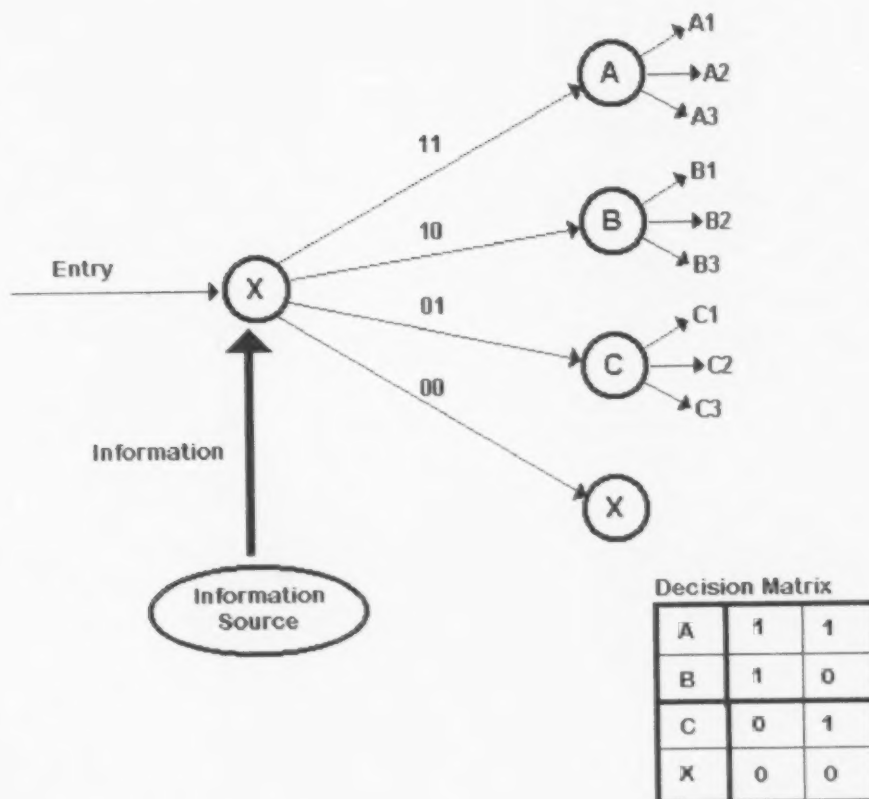


Figure 2: A decision tree and decision matrix.

Each of the pathways shown in Figure 1 differs from the others as a consequence of the different problem states that are visited during the transition from start to finish. These pathways (and hence, the particular states that are visited) will be determined by the decisions that are made during problem-solving. Naturally, the actual results of decision-making are extremely dependent upon the nature and the quality of the information that is available at the points (in Figure 1) at which decisions have to be made. Because of its importance in problem-solving, information has often been called the 'fuel' of decision-making.

The importance of information in decision-making is reflected in Figure 2 which shows a typical 'decision tree' corresponding to the choices of action that can be taken at a typical node during the course of solving a particular problem. Normally, each of the problem states within a given problem space like that shown in Figure 1 will correspond to a node in a graph similar to that depicted in Figure 2.

As was mentioned above, the pathway through a problem space will depend upon the quality and amount of information that is available to the decision makers who are responsible for steering problem-solving activity. Unfortunately, today there is so much information available in the world that no one individual can ever hope to acquire, access, process and use all of it. This plethora of knowledge that is now available has arisen for two basic reasons. First, the never-ending progress that is being made within all areas of human endeavour and its relationships to the natural evolution of the societies in which people live; and second, the breakdown and/or penetration of the international boundaries that exist between nations — thereby making the information available in any one country available in many others. Three important implications of having such vast amounts of information to cope with are:

- (1) we must aim to make it available in electronic form in 'machine independent' ways (Ref 7);
- (2) more effective, automated mechanisms for processing it must be made available;
- (3) easy to use, location independent tools and techniques for accessing this electronic information must be provided.

It is this latter issue which is addressed in this paper. Particular attention is given to the provision of facilities to support 'information on demand': that is, making information available to those who need it when they need it, wherever they happen to be.

2. Types of information

From a problem-solving perspective, the breakdown of the national and international barriers and boundaries referred to above has many important consequences. Undoubtedly, two of the most pressing of these are (1) the growing global mobility of people (such as sales and marketing personnel and technical consultants) and the equipment (or support tools) to which they must have access in order to perform their jobs effectively; and (2) the need to develop appropriate information support services (both software- and hardware-based) that are location independent. Obviously, as individuals move from one location to another the information that each has available will vary considerably. The ease with which this information can be accessed will depend upon the format that it is in, the geographical location of the individual concerned and the types of equipment that are needed for accessing the available information.

In order to provide a basis for the discussion that is presented in the following sections of the paper it is necessary to classify information into four generic types which are referred to by the names: global, isolated, local and mobile. The relationship between these is illustrated schematically in Figure 3. This diagram is intended to reflect how the information available to a given decision maker varies as he or she moves from one particular geographical location (or work setting) to another $\{L_1, L_2, \dots, L_i\}$. Thus, at any given instant in time (t_i) a specified user who is located at a location (L_i) will have available particular measures of global information (g_i) , mobile information (m_i) and local information (l_i) . In addition, it is likely that at the instant in time being considered (t_i) a substantial volume of potentially useful information will be unavailable (u_i) . This is referred to as 'isolated information' in Figure 3.

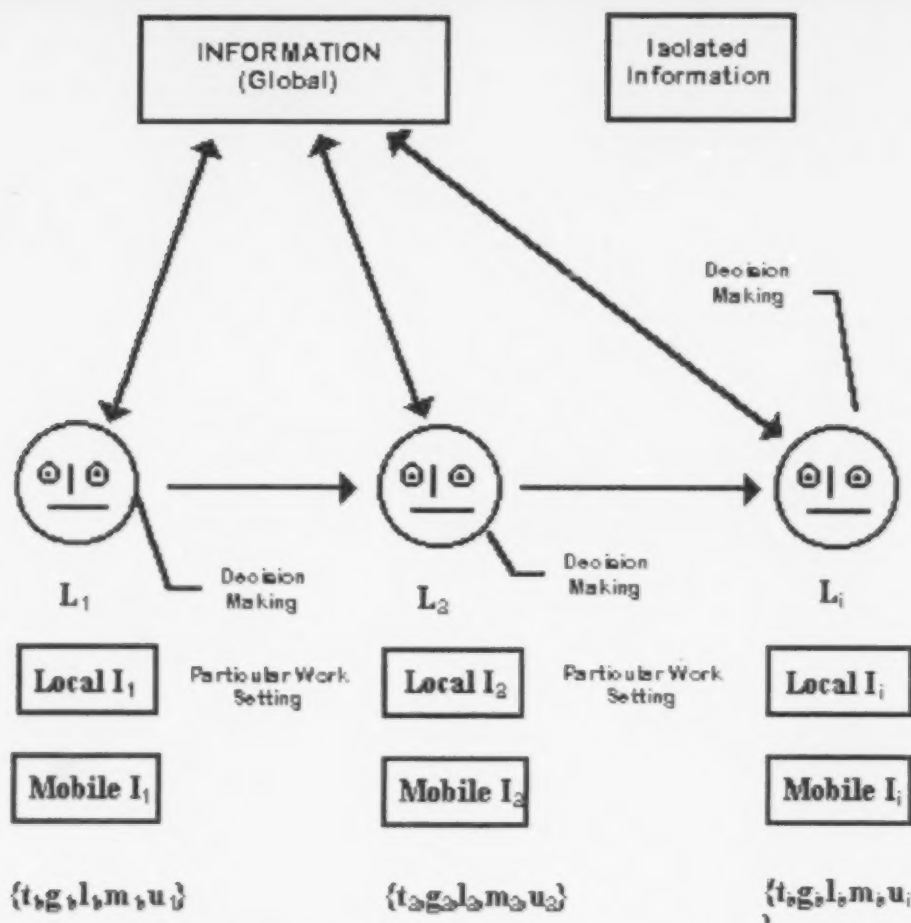


Figure 3: Relationship between local, global, mobile and isolated information.

Global information is information which is easily and widely accessible on a world-wide basis through electronic communication systems. These may be private or public networks and/or broadcasting systems of various sorts. Typical examples of global information are files of material that can be obtained from the Internet (using software such as FTP, for example) or downloaded from the World Wide Web. Obviously, global information may be either private (and possibly encrypted for use only within a closed user group) or publicly available for general consumption.

As its name suggests, isolated information is information that cannot be accessed directly from a particular work setting. Such information can arise from three basic types of situation: (1) information that cannot be accessed because it is not in electronic form; (2) information that is in electronic form but cannot be accessed because of connectivity problems (these may be transient, temporary or permanent); and (3) cases in which information is available online but cannot be accessed because of restricted access rights (that may be imposed for financial, political, security or ethical reasons). Sometimes it may be possible to access certain types of isolated information using special techniques such as 'fax back', or by requesting a 'media mount' operation to be performed by personnel located 'back at base'.

Local information refers to material that is available to a decision maker at the location where he/she happens to be working. For example, this might be a public or private library, or a workstation in a client's office. In this latter case, the computer equipment might provide access to an associated digital electronic library in the form of a collection of one or more CD-ROM discs, or an in-house 'intranet' facility (Refs 6, 12). This type of information is 'fixed' at the location where particular types of job-related decisions have to be made.

Mobile information corresponds to material which physically moves about with a decision maker as he or she travels around from one location to another. It might exist as material that is embedded within the internal storage facilities provided by a portable notebook computer or a palmtop computing facility. Alternatively, it might consist of data that is contained on floppy disks, portable hard disks, solid state disks or CD-ROM discs that are carried around with the computer system. Obviously, certain types of information that might normally be regarded as isolated (Figure 3) could be converted into mobile information by copying it across to a portable medium —

provided it exists in a digital format. The particular subset to be cross-loaded would obviously depend upon the problems to be solved at the locations that are to be visited.

One of the major objectives underlying the philosophy of information on demand is that we should consider ways of (1) minimising the amount of unavailable/isolated information and (2) maximising the volume of globally available information — bearing in mind the limitations and/or possibilities offered by local and mobile material. Some of the mechanisms by which this may be achieved are discussed in the following section.

3. Using portable computing equipment

As was discussed in the previous section, there is a growing demand for the provision of tools, techniques and facilities to make available information on demand in order to support decision-making and problem-solving processes that are not tied to a particular geographical location. The basic types of resource needed to realise this approach to information provision are illustrated schematically in Figure 4.

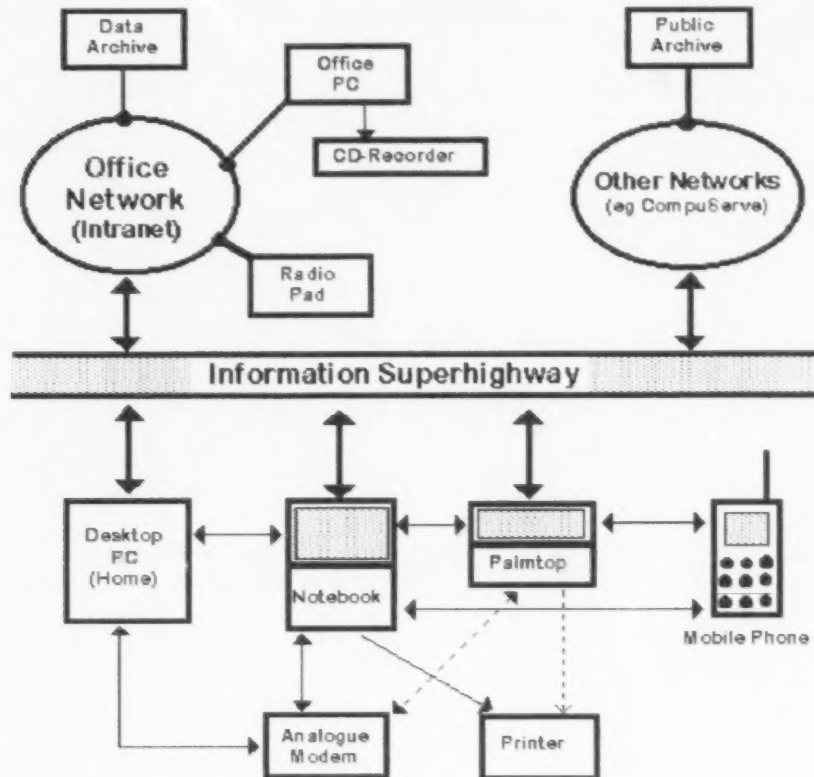


Figure 4: Technology to support information on demand.

In this diagram the important 'integrating' feature is the communication pipe that has been labelled 'information superhighway'. Within this paper this latter term is used in a generic way to refer to all the logical and physical communications infrastructure that is necessary to enable the various computing resources shown in the diagram to be interconnected, in order to facilitate remote information exchange in a global way.

The computing equipment that is illustrated within Figure 4 falls into two basic categories: (1) that which is located at fixed locations (such as the data archives and the office/home-based personal desktop computers); and (2) that which is portable and can be used at different geographical locations — depending upon the types of task to be executed and the nature of the problems that have to be solved. The portable computing equipment includes devices such as notebook and palmtop computers (and their associated peripherals), and digital hand-held portable telephone equipment. The important features of each of these types of resource and the ways in which they can be interconnected to facilitate information on demand are briefly discussed in the remaining parts of this section.

3.1. Notebook computers

Undoubtedly, one of the most popular types of portable equipment is the notebook computer. These come in a variety of different types but their underlying architecture and facilities are usually very similar (Ref 1). Their main features include an A4-size QWERTY keyboard, a high resolution LCD colour panel, a pointing device (mouse or tracker-ball), a large capacity hard disk (usually exchangeable), a CD-ROM drive (either built-in or external), embedded facilities for sound recording and reproduction, a floppy disk unit and a series of interface ports to enable connections to external devices to be made. The external ports usually include a serial communications port, a parallel printer port, a connection for an external CRT monitor and a PCMCIA facility.

PCMCIA is an acronym for Personal Computer Memory Card International Association (Ref 13); cards meeting this specification (Types I, II or III) have become an extremely popular way of expanding the facilities available on a notebook PC. For example, a typical PCMCIA port will allow a notebook to accommodate additional hard disks (such as the Callunacard), memory cards, serial ports (RS-232), fax/modem and data cards for connecting to analogue and digital (cellular) telephone facilities, SCSI devices (SCSI is an acronym for Small Computer Systems Interface) and various types of network card for attaching directly to local area networks (such as Ethernet, Token-Ring, and so on). A useful introduction to the PCMCIA standard and a summary of the different types of product available has recently been published by DIP Systems (Ref 5). The importance of the PCMCIA standard is reflected by the growing amount of online information relating to it — a recent Internet search revealed the presence of 9117 documents indexed on the word 'pcmcia'.

One of the attractive features of the PCMCIA equipment is the ease with which it allows the transfer of information from a desktop PC onto a portable hard disk, which can subsequently be plugged into a notebook computer. Mobile information (see Figure 3) can thus be carried on one or more PCMCIA ATA compatible hard disks. If more substantial volumes of information are to be transported then it is possible to make use of recordable CD ('gold') discs, using equipment that is attached to one of the resident desktop PCs. Also, as is discussed below, by using appropriate PCMCIA data and modem cards, it is possible to connect a notebook computer up to a digital mobile telephone or a conventional telephone network, thereby facilitating access to global information at various remote host sites.

3.2. Palmtop computers

Over the last few years there has been a tremendous growth in the popularity of hand-held computing devices. There has also been a substantial increase in the range of equipment that is available, varying in complexity from simple digital timers through dictionaries, language pronunciation aids and spell-checkers to personal organisers and fully fledged computing systems. Amongst the earliest of the 'more sophisticated' palmtop computers are included the Apple Newton Messagepad, the Amstrad PenPad PDA 600 and Acorn's Pocket Computer (Versions I and II). Systems like the Newton and the PenPad use a stylus to facilitate hand-written input of commands and data whereas the Pocket Computer uses a QWERTY keypad. Since the early introduction of these 'personal digital assistants' (or PDAs) there have been many rapid developments, and there is now a considerable range of equipment to choose from. Some of the latest examples of PDAs have recently been reviewed and compared by Rockman (Ref 12).

Palmtop systems are usually much smaller than notebook computers. Some typical dimensions (in mm) of a palmtop are 166 x 86 x 22 — compared with the corresponding measurements for a notebook which are 280 x 218 x 50. PDAs are also much lighter — typically, 275 gm compared with 2950 gm (a factor of ten less!). The essential hardware features of a PDA include a 'miniature' monochrome LCD display panel measuring about 134 mm (diagonal) and an augmented QWERTY keyboard that usually contains a range of extra keys for activating special functions and/or applications (such as a word-processor, spreadsheet, database, organiser, almanac, sound recorder, fax facility, file manager, calculator, 'comms' package, and so on). The screen and keypad are usually organised spatially in a 'clamshell' like arrangement — very similar to a notebook PC but on a much smaller scale.

As well as its internal memory, a palmtop PC will usually have one or more built-in data card slots and/or disk drives to facilitate the input and output of information using exchangeable solid state disks (both dynamic RAM and flash disks can usually be obtained). The capacity of add-on storage devices varies considerably depending upon the type of peripheral and the particular palmtop computer involved.

As is the case with notebook PCs, most palmtop systems also have some form of interface port. These can

be used to connect the computer up to other computers (see Figure 4) or to devices such as a printer or a modem. Just like notebook PCs, one of the commonest types of interface port is that based on the PCMCIA standard — however, a number of manufacturers employ their own proprietary standards for interface ports.

3.3. Making connections

As can be seen from Figure 4, there are various ways in which the mobile computing elements described above can be interconnected. The two major types of interconnection are those which enable (1) the exchange of local information (between a palmtop, a notebook and, possibly, a desktop system); and (2) the remote connection of a notebook and/or a palmtop PC to remote computers and data archives. These connections can be of either a permanent or a temporary nature. As far as the mobile equipment is concerned, virtually all of the connections will be of a temporary nature.

The exchange of local information can usually be achieved through the use of appropriate serial communication links and suitable client/server software that is installed on the two machines that are to be involved in data transfer. For making remote connections a suitable modem, data card or radio-pad is needed. Undoubtedly, the two most popular ways of making these remote connections employ either analogue (via a modem) or digital cellular networks (via a data card) — although in some cases radio data networks can be used (such as Paknet). Each approach has its advantages and disadvantages in terms of accessibility, cost and the information transfer services that are available. Some of these issues will be discussed in more detail in the case study that is presented later in the paper. Obviously, in order to access global information remotely using mobile equipment some form of server has to be set up 'back at base'. Some ways of doing this are briefly discussed in the following section.

3.4. Setting up information servers

One of the easiest ways of creating a 'base station' server is to use a desktop PC and a modem that is set up in 'call answer' mode. If a conventional public telephone line is then used to dial the modem, remote callers can gain access to information that is held on the server — just like a password protected bulletin-board system. Alternatively, if some sort of Internet connection (or similar network facility) is available then it is possible to use programs like FTP and telnet to gain access to and retrieve global information that is resident on a remote host.

Another important type of server facility for providing access to global information can be built using the World Wide Web (WWW) system. This can be used as a foundation for building quite sophisticated information structures which users can access remotely. There are three fairly simple approaches to mounting information on a Web server, depending upon the amount of information to be made available and the kinds of computing equipment available. The simplest approach is to build a series of pages and mount them on an office (or institutional) server that is located 'back at base': these pages can then be accessed directly with a suitable browser using an appropriate URL address. Alternatively, it may be possible to rent space on someone else's WWW server: this is a useful approach when relatively small volumes of information are involved and the costs of renting are attractive. The other approach is to set up a personal server running on a totally independent machine using software such as Spinnaker's Web Server (Ref 16). Obviously, each of these approaches has its attractions and limitations. In the case study described in the following section each of the three approaches has been used to provide access to information in a global way.

4. A case study — the Psion 3a

For many years I have been an avid user of notebook computers for accessing information while away from my home or office base. However, there are a number of ergonomic and technical limitations associated with using this type of equipment and so I recently explored the possibility of using a palmtop computer to overcome these. The main limitations of notebook computers as mobile resources are: (1) they tend to be bulky and quite heavy to carry (usually requiring some sort of shoulder-bag for transportation); (2) they have a sizeable 'footprint' and require a considerable amount of desktop working area for comfortable use; (3) they require quite a substantial amount of power and so there is only a limited working period available when using batteries; and (4) depending upon the initialisation routines that are invoked on power-up, there can be a substantial delay in getting at required information. Palmtop computers usually do not suffer from any of these limitations. Indeed, one of the attractive features of the Psion 3a (see below) is that it is possible to get access to its associated mobile information using just three key-presses (provided the system is not password protected). These involve: (1) switching on; (2) selecting the required application (say, a database or a spreadsheet); and (3) activating the application. All together, these operations take less than 10 seconds!

In the case study that is presented in the remainder of this section, the mobile computing equipment that has been used is based on the following three items: (1) a DECpc 425SLC/e Premium notebook PC with a PCMCIA port; (2) a Psion 3a palmtop computer with a full range of accessories (Ref 9); and (3) a Nokia 2110i mobile telephone with an accompanying PCMCIA compatible cellular data card (Ref 8). The case study provides an outline description of how this equipment can be used to provide an approach to providing real information on demand.

The Psion 3a is available in a range of different internal memory sizes from 256 Kbytes through to 2 Mbytes. The models with larger memory capacities usually have extra built-in software provided (such as a spell-checker

and a thesaurus). The computer has two peripheral ports for the insertion of solid state disks (SSDs); these vary in capacity from 128 Kbytes up to 4 Mbytes. The disks can be of two basic types: flash and RAM. Information held on RAM can be updated dynamically but that on flash disks cannot (other than by copying a new version of a file to its host disk). The system is powered by two AA batteries (typically, these provide up to 60 hours of use) and/or by a mains adapter. There is a single multi-purpose 'communications' port available which enables various types of system connectivity to be achieved (as is illustrated schematically in Figure 4).

The essential connectivity for the realisation of information on demand can be achieved using four basic types of inter-connecting cable. The simplest of these allows a Psion to be connected directly to a printer to facilitate the output of hardcopy. The second type of cable (the '3Link') is a serial RS-232 link which enables the palmtop to be linked to another computer (either a desktop or a notebook). This allows files of local information to be transferred between a Psion and other computers (at speeds up to 19 200 baud); it is also useful for backup purposes, enabling all the Psion's memory stores (both internal and SSDs) to be backed up onto a notebook or desktop computer. Two useful packages for this purpose are *RCOM* (for Microsoft DOS users) and *PsiWin* — the latter is able to perform automatic file conversion as they are moved/copied from a Psion to a computer running Microsoft's Windows (and vice versa).

From the perspective of accessing global information and communicating electronically with other colleagues, two of the most useful add-on cables are the *3Fax* fax/modem accessory (Ref 10) and the serial link that can be used to connect a Psion up to a Nokia 2110 digital mobile phone (Ref 11). The utility of each of these, from the perspective of providing information on demand, is briefly discussed below.

Global information can easily and rapidly be accessed using a Psion since, as well as facilitating the transmission of faxes, the analogue *3Fax* modem allows the palmtop to be used as a mobile remote terminal from any location within the UK that makes available a standard British Telecom phone socket. In order to do this when travelling abroad, an appropriate set of modem adapter plugs usually has to be used to connect the Psion up to any particular national telephone network (Ref 15). Of course, to enable a Psion to act as a mobile terminal it is also necessary to install a suitable software communications package, such as the VT100 emulator from Widget Software (Ref 18). This allows remote computers to be accessed in the same way that a conventional terminal can be used: files can be uploaded and downloaded; e-mail can be sent and received via a remote host; and it is possible to access the Internet and World Wide Web (using the Lynx character-based browser). In situations where a telephone socket (or an ordinary phone) is not available, the Nokia 2110 mobile phone and a notebook computer can be used to download/upload files that, in turn, can be ported to/from a Psion via its serial link. Unfortunately, there is no terminal software presently available to enable a Psion to be used as a portable remote terminal using a mobile phone and a cellular data network. However, there is a simple 'telenote' facility that enables the Psion to send and receive short text messages via a mobile phone.

The 'telenote' facility is one of a number of SMS (Short Message Service) facilities that digital cellular networks now make available (Refs 11, 17). Telenotes, which may be up to 160 characters in length, can also be sent from an ordinary desktop or notebook computer using dial-up facilities (through an analogue modem). They may be sent to an individual or a nominated group of people, who can receive them on suitable digital mobile phones. If any particular targeted telephone is not online when a message is sent, it can be buffered and relayed onwards to the phone when it eventually comes online. Facilities of this sort are very useful for passing out 'snippets' of important information dynamically to decision makers and delivering urgent messages to colleagues.

Using the types of equipment described in this case study it is possible to implement a variety of physical mechanisms to facilitate the realisation of 'information on demand'. Unfortunately, having access to vast volumes of electronic information brings with it the problems of understanding its structure and knowing how to navigate through it in order to locate particular items of interest. In order to overcome these problems we have been using a metaphorical framework based upon the use of book, bookshelf and library metaphors. The ways in which these metaphors have been implemented are described in more detail elsewhere (Refs 2, 6, 14). Another important direction that our research is currently exploring involves the implementation of a 'virtual office' environment in which desktop, filing cabinet, telephone, fax and mail metaphors are integrated into a uniform and consistent metaphorical framework for the execution of office procedures (Ref 3). Obviously, such an office does not exist as 'bricks and mortar'; it only exists electronically within the shared cyberspace environments that can be made available through the use of distributed and dynamically re-configurable computer networks.

5. Conclusion

Mobile computing technologies of the type described in this paper have paved the way for the true realisation of information on demand: that is, the availability of sought-after information as, when and where it is needed. Although there are still some problems to be overcome because developments in this area are so rapid, the barriers that these pose will soon disappear (Ref 4). Obviously, this new technology brings with it important new opportunities for the realisation of many novel approaches to the realisation of online information. These can embed both well-established and forward-looking metaphorical frameworks that are able to make available new types of online environment which can provide better methods for organising information, and give easier access to it. Such frameworks will naturally place great (often unanticipated) demands on system designers. Furthermore, because mobile computing (and information on demand techniques) can be used to solve such a wide variety of problems, it is imperative that a dynamic and flexible system architecture can be developed and maintained. This is easily achieved through 'hot swapping' of peripherals on an 'on-demand'

basis in order to achieve required system functionality as and when it is needed. Naturally, ongoing developments in the field of PCMCIA interfaces can be used to meet the requirements outlined above by allowing users to mix and swap the various devices that are attached to their mobile computing equipment.

Professor Philip Barker
Human-Computer Interaction Laboratory
School of Computing and Mathematics
University of Teesside
Middlesbrough
TS1 3BA
UK
Tel: +44 (0)1642 342660
Fax: +44 (0)1642 342067
Mobile: +44 (0)468 587 520
E-mail: Philip.Barker@tees.ac.uk
WWW: <http://www.telesys.com/pbarker>

References

- [1] Armbrust, D. (1996) Pentium 90 Mhz notebooks — 15 multimedia notebooks, *Personal Computer World*, **19** (3), 103–129.
- [2] Barker, P.G. (1992) Electronic books and libraries of the future, *The Electronic Library*, **10** (3), 139–149.
- [3] Barker, P.G. (1996) *Human-Computer Interaction within a Virtual Office Environment*, Working Paper, Interactive Systems Research Group, University of Teesside, Middlesbrough, Cleveland, UK.
- [4] Byte(1996) Special report on mobile computing, *Byte*, **21** (5), 83–106.
- [5] DIP Systems (1995) PCMCIA Product Directory, Albury, Surrey, UK.
- [6] *IEEE Computer* (1996) Special edition on digital electronic libraries, **29** (5).
- [7] Maunder, C. (1994) Documentation on tap, *IEEE Spectrum*, **31** (9), 52–56.
- [8] Nokia (1995) *Cellular Data Card DTP-2 for GSM and DCS 1800 (User Guide)*, Nokia Mobile Phones, Nokia Corporation, Helsinki, Finland.
- [9] Psion (1994) *Psion 3a User Guide*, Psion plc, London.
- [10] Psion (1994) *Psion 3Fax User Guide*, Psion plc, London.
- [11] Psion (1995) *Telenote Link for Vodafone Digital Networks (User Guide)*, Psion plc, London.
- [12] Rockman, S. (1966) Personal digital assistants, *Personal Computer World*, **19** (4), 124–132.
- [13] SystemSoft (1994) *CARDSoft+ User's Guide*, SystemSoft Inc., Natick, MA, USA.
- [14] Tan, C.M. (1996) *Hypermedia Electronic Books*, Draft PhD Thesis, School of Computing and Mathematics, Middlesbrough, Cleveland, UK.
- [15] TeleAdapt (1995) *TeleDapters and TelePaks*, TeleAdapt Ltd, The Technology Park, Colindeep Lane, London, NW9 6TA.
- [16] Telesystems (1996) *Spinnaker Web Server*, Telesystems Ltd, Prestwood, Bucks, HP16 0ND, UK.
- [17] Vodafone (1996) *Vodafone Telenote Services — A Range of Short Message Services*, Vodata Ltd, Dower House, London Road, Newbury, Berkshire, RG14 1JD.
- [18] Widget (1994) *VT100 Emulator for Psion Series 3a (User Manual)*, Widget Software Limited, Knebworth, Herts, SG3 6EX.

Electronic references

Barker: <http://www.telesys.com/pbarker/online96a.htm>

Nokia: <http://www.nokia.com>

PCMCIA: <http://www.systemsoft.com>

PCMCIA: <http://www.systemsoft.com/prodserve/cardsoft.htm>

TeleAdapt: <http://www.teleadapt.com>

Telesystems: <http://www.telesystems.co.uk> widget; <http://www.widget.co.uk> vodafone; <http://www.vodafone.co.uk>



New directions in the delivery of business news

Martin White

TFPL Ltd., UK

Abstract: *The provision of profiled business news services is one of the high growth sectors in the US online market. Some of the leading US-owned vendors are profiled, as well as emerging service operators in Europe. Some of the issues influencing the rate of adoption of these services in Europe are discussed, including database content, profiling procedures, billing and marketing.*

1. Introduction

After a period of quite slow growth companies that provide business news to the manager's desktop PC are now amongst the most rapidly growing companies in the US information industry. Well established companies such as Desktop Data and Individual Inc. are now being joined by a number of new services and by enhanced versions of existing services, as well as services based entirely on WWW technology. The objective of this paper is to look at the trends in this sector of the information market, and in particular to look at the market prospects in the UK and the rest of Europe. It is not intended to provide detailed information on all the services currently available.

2. The market for news

In order to give some indication of the demand for news it is interesting to look at the circulation details of newspapers throughout Europe. In Table 1 (overleaf) the first column gives the number of WWW pages, the second column the number of WWW pages per head of population, and the third column the readership of national and regional newspapers per 1000 head of population.

Although the European market for business news services is still in its infancy, this table might indicate the fact that the Scandinavian countries could be early adopters of business news services; and indeed one of the pioneer services, Esmerk, is a Finnish company.

Even in the USA the current market for profiled news services is quite small, perhaps around \$50 million, and our research in the UK would put the market at the level of \$15 million. To date the penetration of these services in the rest of Europe has been very low, with Desktop Data for example reporting only 40 sites in Europe in early 1996.

Table 1

Country	WWW pages	W/p	Readers/1000p
Luxembourg	30,000	13	N/A
Finland	231,000	22	1860
Sweden	313,000	28	1658
Netherlands	298,000	51	1055
Denmark	89,000	59	1056
Austria	115,000	69	1011
Ireland	50,000	71	944
UK	535,000	109	1143
Belgium	89,000	117	748
Germany	564,000	144	971
Italy	251,000	227	477
France	237,000	240	802
Portugal	40,000	242	732
Spain	101,000	384	587
Greece	NA	NA	401

The information on WWW pages is taken from Wired, June 1996, and the newspaper circulation information is derived from a number of industry sources.

3. Market segmentation

Within the broad sector of business news services there are a wide range of services providing access to business news. These include:

- services providing the full text of press releases, notably PR Newswire;
- full-text retrieval services offering archival retrieval of newspapers and other business news sources, but usually with some delay in making the material available, such as NEXIS;
- general newswire services, for example Associated Press and Reuters;
- profiled news services, such as First! from Individual Inc.

The market sector attracting most attention is currently that for customised news services, in which users are able to select, using a range of profiling approaches, only those news stories of direct relevance to them. This profile is run against new stories on a regular basis, often twice a day. Among the benefits that these services seek to offer are:

- eliminating the need for a user to scan through a multiplicity of sources;
- providing information from multiple sources in a common format;
- delivering the content on a timely basis;
- providing summaries of news items;
- providing interfaces for internal transmission of information through an intranet;
- providing archival retrieval as well as current awareness.

Some of these act only as integrators of newsfeeds and provide fairly limited profiling, whereas others provide very substantial levels of editing and indexing, and this is of course reflected in the subscription price.

4. Two pioneers

The two companies that were arguably the first into this market sector were Desktop Data and Individual Inc., both of which were set up in 1988 but did not really emerge onto the market until 1990. The approaches taken by these two companies are very different. Desktop Data's principal product is NewsEDGE, which delivers an

integrated news service based on some 600 sources onto a LAN server on the subscribers premises. One of the attractions of the pricing of the service is that the cost per subscriber decreases substantially as the number of subscribers increases.

Individual Inc. adopted the content analysis methodology developed by the late Gerry Salton at Cornell University. SMART, System for the Manipulation and Retrieval of Text, is a highly sophisticated content-sensitive software suite which can be 'tuned' to improve the relevancy of the profiles as the needs of the user change. First! is a corporate subscription service but over the last two years the company has launched a number of complementary services (HeadsUp, NewsPage and iNews) which are targeted at the smaller company, or the individual subscriber in a larger company, and which provide summaries (usually a few lines in length) of news items. The company offers a wide range of delivery channels, the most popular being fax, e-mail and Lotus Notes/Microsoft Exchange. In 1995 65% of delivery was through e-mail, though the balance is now changing in favour of intranet delivery.

Although both these companies are growing very rapidly in terms of revenues, profits are still a long way off. For example revenues for the first quarter of 1996 for Individual Inc. were just over \$5 million, a 46% increase on the same period in 1995. However the company had a net loss of \$2.5 million for the quarter, compared to \$1.7 million for first quarter of 1995. The company had an Initial Public Offering in March 1996 which realised over \$35 million, and one of the minority shareholders is Microsoft, so the company does not need to be too concerned about its losses just yet.

5. Other business news services

There are many other companies offering business news services with some degree of profiling. In the USA Dow Jones has recently relaunched its Dow Jones News/Retrieval service, though using a Windows interface rather than adopting a WWW browser. In the UK both the *Financial Times* and Reuters are offering profiled services. The FT launched FT Discovery at the end of 1995 in order to be competitive with Reuters Business Briefing and Business Alert. Neither of these services has a WWW interface at present. Reuters has a very wide range of sources and very considerable technical and journalistic resources, and can play a waiting game as some of the new players enter the market.

One of these is MAID, which provides an intuitive interface for its Profound service and which has embraced Internet delivery in a very comprehensive fashion. One of the puzzles for companies in this sector is how to take advantage of Internet/intranet delivery technologies without having to rebuild their existing user interfaces. The Hoover service developed by Sandpoint has been purchased by Information Access Corp., which has a very wide range of business information sources and will no doubt be continuing to invest in the service.

One of the interesting new entrants to the US market is IBM, with infoSage. This offers very comprehensive profiling capabilities and was well received at its launch in early 1996. IBM has the infrastructure to bring the service to Europe but there has to be a concern about the level of commitment of the company to the provision of information services. As well as infoSage IBM offers infoMarket and the Lotus Newsstand services. Just how all these services will be positioned with respect to each other and the marketplace remains to be seen.

6. NewsEyeQ

The NewsEyeQ service, launched in May 1996 in the UK by News International, deserves particular attention. The technology is based on the EyeQ profiler developed by Data Times in the USA (and NewsEyeQ is a joint venture between News International and Data Times), together with Personal Library Software as the retrieval software. The investment that News International has put into the service is very substantial, with a very strong field sales team. This is the first service of its type in the UK to adopt a pay-per-view pricing model, and it will be very interesting to see how this approach works. In the UK market the main competition will come from FT Discovery and Reuters Business Briefing, as well as MAID, but neither of these currently offers pay-per-view pricing or permits credit card billing of the monthly subscription.

7. Broadcast news services

Another fast developing area of business news delivery is that of the provision of television news on a 24 hour basis. This was pioneered by CNN but there are two important new services due to be launched in 1996. Microsoft and NBC have set up a joint venture into which they are proposing to invest \$620 million over the next five years. Initially this will be broadcast television, but an extension into interactive news services is also proposed. As has been mentioned above, Microsoft has a stake in Individual and may be using this as a means of developing search and profiling techniques for the new service.

News Corporation, the parent of News International, is also moving into this sector in the USA based on the Fox Television network, and is marketing this very aggressively to cable operators.

The importance of these services is twofold. First they will raise the visibility and importance of business news. However, possibly more important is the possibility of using the technology developed by Intel to link a teletext-

type signal broadcast with the news stories to link into WWW sites. In a similar approach LEXIS/NEXIS is working with CNN on a link with the daily CNN broadcasts to provide research support on stories cited in the broadcasts.

8. Market development issues

The basic commercial issue is how to be able to deliver highly relevant news at a price that represents value for money. With most other online services the decisions on relevance are left to the end user, and because each search will probably be for a different subject it is not easy for a user to judge how good the service is at finding relevant articles.

The situation is very different with profiled news services. Each day, or even several times a day, the user will look at the results of the profiling process and be able to judge quite easily which of the (typically) twenty stories delivered that day are in fact relevant and timely. The user will also be reading newspapers and industry journals, and be talking to colleagues and customers, and does not want to be in the position of missing an important news item. Keeping the subscribers satisfied in this market sector is going to remain a very considerable challenge and will entail continuing attention to content, profiling, delivery platform and pricing. The marketing challenges are also considerable.

9. Content

Judging these services on content is also a considerable challenge to subscribers. Some of the services, such as Reuters Business Briefing, have over 4000 sources but many of these are taken from other news services such as Information Access Corporation. As a result the rate at which the service is updated is often difficult to ascertain other than for feeds from daily newspapers. Some newspaper publishers are also (not surprisingly) protective as to through which services they will allow their feeds to be accessed. For example, NewsEyeQ does not have the *Financial Times* — a result probably of the ownership of NewsEyeQ by News International, a major competitor of the FT. A good example of the complexities of the current situation is given by Associated Press which has four newswire feeds, including AP Online. At present these are available through eight distributors, but in some cases there is a time embargo and in others the material is rewritten.

For US services important decisions have to be made about extending the coverage of their services towards European news. There are a number of issues here. The first is one of language, not only because their current subscribers are unlikely to want to read *Le Monde* in French but also because of the need to adapt their profiling procedures. As any visitor from Europe to the USA will know, the level of coverage of non-US events (especially where there is no US interest, such as the 'beef scare') is small, so adding a substantial amount of even English-language European content is unlikely to be a major selling point with US customers.

The second is the wide range of viewpoints that European newspapers give their readers. This is especially the case in the UK, where it is important to read the slant to a story as reported by the *Financial Times*, the *Times* and the European edition of the *Wall Street Journal*. Should the stories be de-duplicated? Related to this is the fact that a number of journals, with *Business Week* and the *WSJ* being good examples, have US and international editions which have different stories, sometimes in differently-dated issues.

10. Profiling

Once the content parameters have been established the profiling needs to match both the content and the needs of users. This is where differences in the use of English across the Atlantic start to be significant. Gas and petrol are but one example. In an online search the database provider may well have built a list of synonyms and/or the users may themselves be aware of the problem when constructing the search request. If using preset topics as the basis for the profiling then a good understanding of the scope of the topic heading is essential. This in fact may be difficult with some of the complex profiling routines. Although inspection of the results of the profiling can help to remove unwanted articles, it is difficult to recognise what is not present and should be!

Just providing 'see also' links is only part of the problem. Topic lists generated by US companies are often structurally unable to cope with concepts such as governmental structures in Europe. There is also evidence of a significant lack of basic awareness of European industry. On one list of topics (the company must remain confidential), under European Airlines there was an entry for Air UK but not for KLM, SAS or Lufthansa.

11. Delivery platform

The objective of all news services is to ensure that as many managers as possible in a company have access to their services. In this respect the services are reaching what might be regarded as the 'ultimate end user' who may use no other online services. Both Desktop Data and Reuters Business Alert developed LAN server architectures to achieve this objective but now the range of intranet options is increasing rapidly. The technical issues of supporting both HTML and proprietary platforms (such as Lotus Notes and Microsoft Exchange) are less of an

issue than controlling the onward redistribution of the material in the organisation, and making the appropriate royalty arrangements with the content provider.

12. Pricing

The pricing options at present range from very expensive to free. Moreover most of the suppliers (with the notable exception of MAID) have such complex pricing strategies that making cost comparisons between suppliers is very difficult, if not totally impossible. The formulas are usually based on the number of users, the number of profiles and the number of articles provided, as well as possible additional costs for the delivery of the full text of the article where initially only a summary has been provided.

Giving indications of the typical expenditures by companies on the high-end services such as those from Desktop Data and Individual is always going to be misleading, but expenditures in the range of \$15,000 to \$20,000 are not unusual. The way in which Desktop Data prices its service is especially interesting. The formula is designed to make it very attractive to corporate clients to increase the number of individual users, in order to bring the cost-per-subscriber down substantially. For example a news service can be provided to 5000 users on one server for \$0.15 per day, though this represents an annual expenditure of over \$150,000.

At the other end of the pricing spectrum is Pointcast, an innovative WWW-based service which provides a business news service free of charge, as the service is supported by advertising and sponsorship. With WWW services it is possible to link together complementary services so that more detail on a particular topic can be gained from other databases, with the database owner paying a fee for the link. This form of sponsored service is still a novel one, though the Health Online Service developed in Germany by Burda uses a similar approach.

Of course there are many newspapers providing full text access over the WWW. Some of these tried to move from a free service to a charged service, but with very little success, and most of these sites are now seen as advertising for the printed newspaper. The majority of these sites now require some form of registration, which enables the site owner to develop a mailing list based on readership, something that cannot be obtained through the extensive distribution through newsagents in the UK. The situation in some other European countries is different as there is a substantial subscription-based circulation.

In between these two approaches are the pay-per-view strategies, with a small monthly subscription and a payment per hit or per document that is related to the level of the monthly subscription and the type of document retrieved. This model is being pioneered in the UK by NewsEyeQ, which is also accepting credit/charge card payments in an attempt to attract users that see the annual contracts and high subscription levels of other services as a significant barrier to adoption. Of course this is not just a UK approach, as the model has been used very successfully by Individual Inc. with its HeadsUp service in the USA, though these pay-per-view options have yet to be introduced in Europe.

13. Credit card billing

One of the reasons for this may be the very different attitude of European companies to corporate credit cards, whether they are issued in the company name or are personal cards for which expenses are reimbursed by the company. From research carried out by TFPL it is clear that companies have very substantial reservations to having the monthly subscription fees charged directly to a card, as the paperwork audit to ensure that the expense is actually incurred on behalf of the company is quite considerable and reclaiming Value Added Tax is also very difficult. In this respect the UK is closer to the US attitude, but elsewhere in Europe card billing is a significant problem that has yet to be addressed by most suppliers.

14. Marketing and customer support

The costs of marketing profiling news services is high, as it takes guidance to create the profiles and there will be a continuing need, especially in the early stages of a subscription, for support in revising the profiles or adding to them as the service is adopted more widely in the organisation. This high cost of sale can be accommodated within the subscription levels of services such as Reuters Business Alert, but not for the services targeted at smaller companies or individuals. There is a substantial requirement for the services to develop user interfaces that facilitate the creation and evaluation of profiles, as well as providing solutions to the language problems referred to above.

15. Some future directions

There is no doubt that the business news market will develop rapidly over the next couple of years. It is likely that some of the service providers will start to work with organisations, such as trade associations, to provide members with customised news services. This will reduce the marketing costs and the branding of the organisation will be very important for smaller service providers.

These services may also act in a Trojan horse fashion for other information providers, as they will reach more widely within client organisations than more specialised services and raise the overall awareness of online services. There could be links, for example, that once a user has seen a company cited in a news story, clicking on the company could link the user in the Dun & Bradstreet database, for example.

For many years business information services have been a profitable niche market for a number of suppliers. With new competitors determined to expand the market through innovative product and marketing strategies, this is a sector that requires careful monitoring by both the information industry and information managers.

Acknowledgement

I am grateful to my colleague Mary Coyne for her assistance in the preparation of this paper.

Martin White
TFPL
17-18 Britton Street
London EC1M 5NQ
UK
Tel: +44 (0)171 251 5522
Fax: +44 (0)171 251 8318

Reference

White, M. and M. Coyne (1996) *Business Information in the Intranet Age*, TFPL Ltd., London and New York



Tailored news for everyone — the Reuters business briefing case study

Paul Waddington

Reuters, UK

Abstract: Business users have been connected to computer networks for several years now. Reliable online business information has been available since the 1970s. So why don't we all have desktop access to all the news we could benefit from?

Simple interfaces, good content and flat-rate pricing have gone a long way towards creating services which deliver tailored news for everyone: this paper will include case studies illustrating how Reuters Business Briefing is used now across different job functions in an organisation. The studies will also illustrate the role played by information professionals in this new way of delivering news.

The business world has now woken up to Web technology and the fact that it allows companies easily to realise the potential of their network for sharing, storing and disseminating information. Whereas networks have in the past functioned as little more than devices for sharing e-mail, files and printers, Web technology — in theory at least — makes an organisation's knowledge available to its workers in a format they can easily understand.

Each intranet user is sitting in front of multimedia publications waiting to be filled with content. This paper also looks at ways of integrating news easily into this environment and concludes with a consideration of some of the future possibilities.

Paul Waddington
salesworld@reuters.com



Solving the imaging dilemma for Internet and intranet applications

Brian Meek

Iterated Systems Incorporated, USA

Abstract: Providers of online information with a high visual content face an uncompromising dilemma — how to achieve fast delivery whilst maintaining good picture quality. As storage space or bandwidth increases, the demands placed on it by the user are always two steps ahead. Since limited bandwidth for communicating images looks set to remain a problem for some time, the biggest issue facing graphics developers is how to create a Web site, using today's technology, which remains attractive to all users regardless of their access speed. Their site must be accessible to both ends of the speed spectrum and still be video and image rich for each user.

Keywords: fractal, image, video, compression, Internet, bandwidth, broadband, narrowband

1. Introduction

It has been said that a picture, with its complex mix of shape, form, contrast and colour can paint a thousand words; a belief widely upheld in the multimedia industry. The explosive growth in this area in recent years coupled with the incredible advancements in multimedia technology means that users now expect, and receive, quality imaging on their desktop. But with the incredible shift towards a networked society in the past few years, how then does this translate for online applications? We are living in a networked society and the phenomenal growth of the Internet and the intranet, along with client server and groupware technology, means that most organisations and many individuals receive incredible amounts of information from networks. The dilemma facing application developers in the graphics field therefore is how to recreate the same high quality image level for online applications as they are able to provide on the desktop.

2. Current limitations

As the Internet has grown, the weakness in the network as a delivery mechanism for pixel-based information has become clear to even the most casual network user. The World Wide Web has become the World Wide Wait. Technology developed for the personal computer and its small bandwidths doesn't translate well for imaging in Web sites; it was, after all, not originally developed for this unique environment.

Yesterday's, and to a large extent today's compression technologies have all evolved with an eye toward storage optimisation. They were designed with a mind set that storage was at a premium and bandwidth was unlimited. With the advancement of technologies, it is becoming apparent that the reality is just the opposite. Storage is relatively inexpensive and bandwidth is at a premium and, as storage space or bandwidth increases, the demands placed on it by the user are always two steps ahead. Since limited bandwidth for communicating images looks set to remain a big problem for some time yet, the biggest issue facing graphics developers is how to create Web sites that are attractive to all users regardless of their access speed. Sites must cater for both ends of the speed spectrum and still be video and image rich for each user. Hence the need for data compression to maximise the usage of the existing infrastructure.

3. Data compression technologies

Data compression can be categorised into three groups: 'lossless', 'lossy' and 'visually lossless' technologies.

3.1. Lossless technologies

Lossless compression allows you to retrieve the information that you put into the compressor in exactly the original format. Examples of lossless compression include huffman or arithmetic-based encoding schemes found in products and technologies like PKZIP, Fax Group 4 or LHARC. In addition the GIF format predominant on the Web uses a similar lossless approach. This sort of compression yields fairly shallow compression ratios: up to

about 4:1 depending upon your source material.

3.2. Lossy and visually lossless technologies

By contrast, lossy compression achieves much higher compression ratios by permanently discarding certain information from the original image. Obviously, for ASCII text this can be a quite inappropriate method but when manipulating image data it proves to be extremely effective, since uncompressed digital images contain redundant data that is not necessary in order for the image to be recognisable: take, as an example, a photographic print which is a 'lossy' representation of a real world view. So, clearly there are degrees of 'lossiness' that we are prepared to accept.

Examples of lossy compression techniques include discrete cosine transformation (the maths used for JPEG images), Wavelet and Fractal. With lossy compression technologies, the user chooses the compression ratio based on the amount of data loss that the image will tolerate. This is obviously a totally subjective issue and the nature of the loss varies according to the different algorithms used. At high compression ratios, JPEG and Wavelet tend to produce a more blocky resulting image with pseudo colour being introduced, and these artefacts are exaggerated in printed versions of the image. Because the nature of the fractal mathematical process means that it operates outside of the pixel-based world, highly compressed fractal images degrade more gracefully and print out quite smoothly.

The term 'visually lossless' is used to describe an image compressed using a lossy technique but to a shallow enough compression ratio for the user to be unable to recognise a difference between the compressed and original versions. Again, this is a subjective issue and varies depending on the project for which the images are to be used.

4. The advantages of fractal mathematics

Since its founding in 1987, Iterated Systems, specialist in imaging solutions, has pioneered the application of fractal mathematics for the management of digital images, seeing benefit not just in the high compression ratios achievable but also in the other inherent advantages of turning an image into a fractal format. In 1987 Michael Barnsley, co-founder and now Chief Scientist and Technology Officer at Iterated, discovered the mathematics behind the fractal transform process opening up a whole new era in digital image representation.

A fractal image is an infinitely magnifiable picture that can be produced by a set of instructions and data. Iterated Systems' fractal technology compresses images as recursive images: that is, the fractally compressed images are stored as recursive algorithms — sets of equations with instructions as to how they should be reiterated in order to reproduce the image. The equations describe the image in terms of internal relationships between its components, rather than pixel-by-pixel descriptions of its qualities. The tighter compression rate is possible because only these algorithms are stored, rather than a pixel representation of the image.

4.1. Resolution independence

Because the fractal maths formula seeks out shapes, patterns and colours (not pixels) within the larger picture and how they relate to other images within the picture, the formula for creating a particular graphic always remains the same, regardless of the resolution of the image. Think about it for a moment — an infinitely zoomable image. Applications are truly scaleable to the colour and size resolution of the display device (or printer) being used. Images can be scaled to full screen or zoomed to locate finer detail, while still retaining image quality. Fractal images can also be decompressed to colour depths both higher and lower than the original. This resolution independence ensures that pictorial content displays well on a wide range of hardware and protects the time the developer invested in its creation. By contrast, JPEG compressed images are resolution dependent. Users cannot display the original decompressed image at a higher resolution without pixelation and blockiness. The implications to developers or publishers are that multiple digital images may have to be managed concurrently; and potentially, with electronic imaging input and output technology resolution increasing at a breakneck speed, these resolution dependent images must be rescanned and recompressed to keep up with the changes.

4.2. Content addressability

The set of 'equations' that make up a fractal image describe the relationships between the components of the image, rather than the individual pixels that make up the bitmapped representation. This property makes it possible to search a fractal image (or even a collection of images) for near matches with other visual patterns, i.e. images with similar fractal components, even if the content is rotated, scaled, skewed and coloured. This makes the fractal approach suitable as the basis for identification purposes in applications that require fast searching and image comparison for security, image management, database systems and so on. The implications of such technology are astounding.

4.3. Fast, software-only decompression

Fractal images render to on-screen pictures very quickly. The computationally intensive operations in the compression process do not need to be repeated. The instructions governing the shape and positioning of objects within the image can be performed efficiently and in software alone. In comparison to a JPEG image, the FIF will decode and display faster: across a low bandwidth network connection the disparity is even more discernible.

Online Information 96 Proceedings

4.4. Progressive decompression

One could say that the real value of compression technology is in the delivery of the image. The methods of progressive and controllable decompression of fractal images mean that online developers and their eager viewers can experience a visual display vastly superior to what conventional techniques allow. A 'thumbnail' image appears almost instantly at full quality, which proceeds to grow before the user's eyes into a full resolution image as data is transferred across the network. The surfer can now enjoy full-focus images almost immediately, and is able to continue traversing the Web at maximum speed without waiting for the image to complete its download process.

5. Video compression for online applications

We have already discussed the new genre of online applications appearing on our networks today (World Wide Web or otherwise) embracing commercial business applications, entertainment and information delivery. The primary focus of the paper thus far has been how to provide a solution for the delivery of *images* within this new online environment. But what happens when you introduce video into the equation?

Consider for a moment the number of frames required to transmit just one second of video and it quickly becomes apparent that the issues associated with the transmission of video over the online networks are quite literally 25 times (or 30 times for our transatlantic colleagues) more severe.

5.1. Video compression technologies available today

A whole plethora of video compression techniques are available, each optimised for a wide range of diverse applications. People wanting to achieve video delivery over broadband network connections (typically between 10 and 50 times ISDN) are well served by the MPEG-1 and MPEG-2 techniques that carry high-quality compressed datastreams to the desktop, and potentially the home.

But what about the growing number of people needing a solution for narrowband connections? Using MPEG-1 technology, a minute of video would require a 50 minute download wait via a modem connection. The new MPEG standard for narrowband delivery, MPEG-4, will use quarter common interchange format (QCIF) resolution and low bit rates for teleconferencing applications. Ratification is expected for 1998. But with ISDN connections a luxury to the majority of Internet users today, a video compression solution is needed now that will operate in the narrowband arena on connections ranging from 14.4 kilobits per second up to 128 kilobits per second.

5.2. Realistic expectations of online video

When considering a solution, we must not be too quick to criticise the playback speed and image quality of multimedia video clips without considering the technical side. We must face up to what are achievable goals for the delivery of video over low bandwidth networks and acknowledge that, at present, we are constrained by the network infrastructure and the issues surrounding this. Although higher bandwidth is an emerging reality it is not, in isolation, the solution to our problems.

The immediate goal is not to deliver TV pictures to a network computer or home browser. We all consider ourselves experts in watching TV and, to even an unsophisticated eye, a quarter screen movie playing back at 15fps is not what we are used to seeing. Broadcast quality video is not the goal. What the technology will do, however, is bring a whole new level of interactivity and user choice to online applications. Remote banking and distance learning will be taken to new levels and emerging applications such as home shopping, narrowcasting, news-as-it-breaks and interactive tourist kiosks will all be brought one step closer to achieving their original design and commercial goals.

5.3. Fractal video compression

Fractal video compression is optimised for low bit rate video delivery in online applications and addresses the issues that existing video compression schemes hadn't even considered. It enables you, the publisher, to deliver your message the way you want it to appear, more clearly, and to a wider audience in less time.

If your video message is low-action as with a talking-head clip, filmed in front of a static background with minimal movement, then a compressed video data rate of around 10 kilobits per second may provide perfectly acceptable quality. Take, on the other hand, a high-action video message. Obviously the more action, the more data you require in the compressed video stream to deliver the message clearly — a rock music video, for example, may require over 100 kilobits of data per second to have the desired impact.

Clearly then we need a scaleable technology: one that will work well at both ends of the spectrum. Fractal technology gives us this scaleability with a unique video compression method that exceeds our expectations by providing surprisingly good results at very low data rates, and increasingly higher quality results as more data is allocated to each second of video: ranging from about 14.4 kilobits per second (standard modem), all the way up to 128 kilobits per second (dual ISDN).

6. The products

It is, and has been, Iterated's belief that the delivery of information is key in today's networked world. To support this belief, the company has recently launched a new product range for both still images and video, based on its patented Fractal Image Transform to give surfers, software developers and content providers for the Internet/intranet a solution to the Web imaging dilemma — providing high-quality at low data rates.

These new products are part of a new business strategy, 'fame and fortune'. The aim is to establish fractal technology as the *de facto* standard for images on the Internet/intranet. To assure service and content providers that their fractal images will be widely accessible and viewable, Iterated is aiming for the mass proliferation of fractal technology with free stills and video decoders downloadable from its Internet home page on the World Wide Web. The focus is to provide software-only fractal encoding and decoding products to the mass-market so that the use of fractal technology is as simple and inexpensive as possible.

Iterated has torn down the 'barriers to entry' for the wannabe Web publisher and professional corporate Webmaster alike. Generation of fractal content is executed by shareware products, and delivery uses standard frameworks like Video for Windows and QuickTime. In addition, by embracing emerging Web technologies like ActiveX, Java and LiveMedia your content can be delivered without the additional expense and implications of proprietary Web server software.

7. The technology in use

Developers have long since realised the benefits of using fractal technology to achieve fast delivery and good picture quality for CD-ROM-based applications. It was fractal technology that allowed Microsoft to incorporate 7000 images into its *Encarta* CD-ROM encyclopaedia, and Hutchinson and Grolier soon followed suit for similar projects.

In today's rapidly expanding online world, the potential for fractal technology is limitless. Smartcard developers, desperate to squeeze the highest amount of information into a limited amount of space, are already making good use of fractal technology to store highly-compressed, high-quality images on smartcards which can then be taken online for a variety of security based applications.

The World Wide Web itself is already playing host to a vast number of diverse fractal-enabled sites, and numbers are increasing on an almost daily basis. Fractal compression is also proving itself to be a major force in intranet-type environments, enhancing functionality and increasing data throughput across local and wide area networks.

8. A customisable technology

Iterated will continue to develop the core mathematics to enhance the possible applications of the technology, and will carry on with the customisation and optimisation of its products to sharpen any specific features required by applications in vertical markets such as medical imaging, security, high finance, entertainment and Web broadcasting. Projects are already underway for security cameras and network computers, for example.

As online graphics technology improves, fractal compression technology continues to undergo significant changes. It is worth noting that since the technology was first introduced, compression times have been reduced by more than 90%. This staggering improvement, coupled with industry advances within the online services sector that provide tremendous opportunities in networked information flow and communication, and future discoveries and improvements can only advance fractal technology beyond its current superior performance level and give value to developers and users alike. Value that is measured, not by the illusion of multimedia lavishness, but by how well it helps you the publishers to sell, educate and entertain.

The next time you run across a site that is fractal-powered, remember, your expectations will be exceeded, clearly.

Brian Meek
Iterated Systems Incorporated
3525 Piedmont Road
Seven Piedmont Center, Suite 600
Atlanta
GA 30305-1530
USA
Tel: +1 (404) 264 8000
Fax: +1 (404) 264 8300
E-mail: info@iterated.com
<http://www.iterated.com>

Bibliography

- Davis, B. (1996) Fractals take us beyond pixel limits, *Advanced Imaging*.
Gilbert, K. (1996) Solving the imaging dilemma, *IDPM Journal*.

Online Information 96 Proceedings

Page 282



On the move: mobile information systems

Johan Hjelm

Bonnier Media Lab, Sweden

Abstract: This paper intends to summarise the experiences we have acquired during the One the Move project and to draw on some conclusions on how mobility will impact information service providers. It begins with a definition of mobile information systems and explains the On the Move project. We then look a little deeper into the mobility aspects. Following that is a discussion of the information aspects, especially how the information demanded changes when the user can access it free from constraints in time and space. We then discuss how this information can be obtained by the provider, which existing systems can be used and which new systems and information must be created. The paper ends with a discussion on presentation aspects, as the same information will need to be presented in different ways, depending on the user terminal and the transmission network capabilities.

Keywords: information services, mobility, multimedia, wireless connectivity, information provision/digital publishing, the future of online

1. Introduction

During the course of the On the Move experiment, we have come across several important issues concerning provisioning a mobile information system. Chief among these is that the presentation must adapt to different terminal capabilities and bandwidth, and be coded accordingly; and that information can become geographically dependent and change with the location of the user. To provide information structured in this way will challenge current systems and infrastructures. The bandwidth issue is widely expected to be solved by the advent of UMTS, the Universal Mobile Telephone System. New problems will then arise concerning video in the interactive wireless environment.

2. What is a mobile information system?

In the future, the boundary between the local area network and global connectivity networks will disappear. This is true also for mobile systems (Ref 20). This means that existing information systems will be globally accessible but it also means that users will be able to access information which needs to be geared towards them while they are mobile.

2.1. What is On the Move and how did we get involved?

On the Move is a European research project funded by the European Commission DG XIII through the ACTS programme. This is a research programme within the fourth framework of European research programmes, with the intention of furthering European research in leading-edge technologies (Ref 1). The partners include Ericsson Radio Systems (Sweden), Bonnier Media Lab (Sweden), KTH (Sweden), SICS (Sweden), EED (Germany), Siemens (Germany), RWTH Aachen (Germany), Burda (Germany), Sony Europe (Germany), Tecsi (France), IBM Europe (France), CWC (Singapore), Iona (Ireland) and BT (United Kingdom).

The ACTS programme is also somewhat different from the earlier RACE and RACE II programmes in that its purpose is not basic research and basic technology, but applications of technology. The programme is still pre-competitive, which means that companies that normally are competitors should be able to cooperate.

The goal of On the Move is to develop a 'mobile middleware', a software that intercedes between the computer and the communications device and manages connectivity in the manner best suited to this. The project is geared towards Universal Mobile Telephony System (UMTS), the mobile telephony system of the future. This system will build on existing mobile technology but have a transmission capacity of 2 Mbps per user. In comparison, current mobile telephones have a capacity of 9.6 kbps per user and an ISDN line of 64 kbps per user. To put this in perspective, a text transmission works fine at 2.4 kbps, but a World Wide Web page with pictures tends to be perceived as slow at 28.8 kbps, and a normal video transmission requires at least 256 kbps.

UMTS is expected to be available sometime after the year 2000 (Ref 20) and the work on standardising it is

currently going on. However, it is not very likely that the system will be deployed all at once. Instead there will be islands of GSM and GSM+, a next-generation development of GSM with capacities of up to 64 Mbps per user. There are also other concerns, like security and charging, that need to be addressed. The On the Move middleware will manage this and shield the interaction with the transmission network from the user.

One of the prime applications of mobile systems is expected to be information services and this is where the Bonnier Group comes in. A \$1.25 billion corporation based in Scandinavia, we already have substantial traditional electronic information services (Hoppenstedt-Bonnier in the UK; Affärsdata, Findata, Bonnier Kreditfakta and Nyhetsbyrå—n Direkt in Sweden; Telebørsen in Denmark) as well as extensive Web publishing activities. We were approached by Ericsson to participate in the On the Move project.

Testing a software product is normally either done through a technical verification procedure (in the case of systems software, for instance) or in a user test. The On the Move project will do so with extensive user testing (extensive for the telecommunications business, anyway) (Ref 13).

It is in this context that we have been confronted with the special demands of mobile users and some of the issues that will need to be addressed by future information systems for mobile professionals.

As publishers, our interest lies mainly in the information service. To create an information service that would test the functionality of the underlying communications software, simultaneously giving something to the user, we provided the following:

- a personalised information service, enabling the user to customise his or her presentation;
- adoption of the layout and presentation to the transmission capacity of the medium (varying from 2 Mbps, which enabled video; to 9600 bps, which was at most sufficient for text);
- location dependent information (as the user moves from one city to another, the 'current city' information will have to change);
- multilingual information.

Personalisation is especially seen as a very important feature of future information systems (Refs 12, 16, 17).

2.2. Mobility issues

Often, mobility is confused with moveability. In the early days of mobile telephony this was understandable: it was not possible to use a mobile telephone except in a car. Today the handset is easily carried in the pocket but it is still mainly usable for voice communication, even though the transmission is digital, and thus should be able to handle data communications as well (Ref 20).

Mobility also comes in two basic flavours: the user being mobile and the information being mobile (Ref 2). Basically, in the first case you access existing systems from a mobile device and in the second case you access information that is being adapted to you where you are.

2.2.1. The user/terminal being mobile

The first set of problems in a mobile information service derives from the fact that the terminal is mobile. This implies several things (Ref 11):

- the network capacity is low;
- the terminal itself has several limitations;
- the user can move from an area with good reception to one with bad, or none at all, with very short notice;
- in the case of a news service, the news feed must be kept coming even if the user moves to an area of bad reception;
- if the user is cut off from reception, some mechanism must be provided to handle the information.

We will look a little deeper into each of these points.

- *The network capacity is low.* Universal Mobile Telephony System (UMTS) will provide a capacity of 2 Mbps per user. In comparison, current mobile telephones have a capacity of 9.6 kbps per user and an ISDN line of 64 kbps per user (experimental next-generation GSM systems have been demonstrated with transmission capacities of up to 64 kbps). To put this in perspective, a text transmission works fine at 2.4 kbps, but a Web page with pictures tends to be perceived as slow at 28.8 kbps and a normal video transmission requires at least 256 kbps.

This implies that except for islands of UMTS connectivity, video is out. As anyone who has tried knows, on a normal GSM line ordinary Web services are out too. However, our experiences as publishers tell us that the nicer the packaging, the higher the user acceptance. So while information will have to be provided as text over the GSM line, it still will have to be nicely packaged.

- *The terminal itself has several limitations.* Currently laptops are the mobile terminal of choice. However, they are heavy and battery life is short. It is very likely that information systems of the future will look more like the Nokia 9000 communicator (Refs 10, 16, 18). This in turn implies limitations in user input and screen capacity (the least of these being that it is black-and-white). These limitations must be conceded when presenting information.
- *The user can move from an area with good reception to one with bad, or none at all, with very short notice.* It is unlikely that all of Europe will be covered with UMTS at once (Ref 20). Even today, GSM does

Online Information 96 Proceedings

not cover all of countries like Sweden, Norway and Finland, where penetration of handsets is highest in the world (Ref 15). Also, radio networks are prone to service quality loss, to cite the reflections of TV signals in London's Canary Wharf as an example. Solar radiation can also cause disturbances and it is unlikely that the system will cover all underground traffic systems (subways, tunnels and so on). This means that some mechanism must be built into the system to handle interruptions and quality changes.

- *In the case of a news service, the news feed must be kept coming even if the user moves to an area of bad reception.* This implies that the quality changes will have to affect the presentation of information, probably according to a list of priorities set up by the user. This in turn implies that there must be a personalised handling of the connectivity.
- *If the user is cut off from reception, some mechanism must be provided to handle the information.* Even if the network adapts to quality change, or even handles interruptions of service, it will be meaningless to the user if he is flooded with telegrams when he has been out of the network for a while. This is especially pertinent in cases where he voluntarily switches the receiver off (for example, during a meeting). That, in turn, implies that the information service must handle these cases.

2.2.2. The information being adopted to mobility

Information itself can adapt to the user being mobile. Today there are receivers for the Global Positioning System (GPS) available on a PC card. These can be used to create position information, possibly in conjunction with the GSM network (Ref 9). If this information is fed back automatically to the sender from the receiver, the information presented can change as he or she moves from one place to another. One of the simplest cases would be city information. However, there are several problems with this approach (Ref 11). For instance, what are the boundaries of the current information set? How does the system receive coordinates? What is the resolution of the database and how is this fed to the sender of information? Should this change if the user changes terminal, or quality of service changes? What if he is in a high-speed plane, or a train?

3. The On the Move information system

In the On the Move project we have developed an information system to meet at least some of the requirements stated and implied above.

The information system is divided in two: one German and another Swedish. This is partly due to the initial availability of the information in one language only (see Section 6 below) but also to the fact that the two systems each work in a different test bed: The Swedish system is intended to test bandwidth adaptability in an information stream environment while the German system is intended to test location dependency. The system runs on standard laptops and is based on the Web with Netscape Navigator as a front-end. The software part has been developed by GSI Tecsi of France while the content is obtained from existing Bonnier and Burda systems (Ref 13).

The Swedish part is a fairly simple information system, based on feeds of news telegrams from Nyhetsbyr—n Direkt and the Web page of *Dagens Industri*, our business newspaper, who publish full listings of the Swedish stock exchange (Ref 11). This system, shown in Figures 1 and 2, can be fully customised by the user by feeding in keywords and stock names to create a personal portfolio of news and stock exchange information.

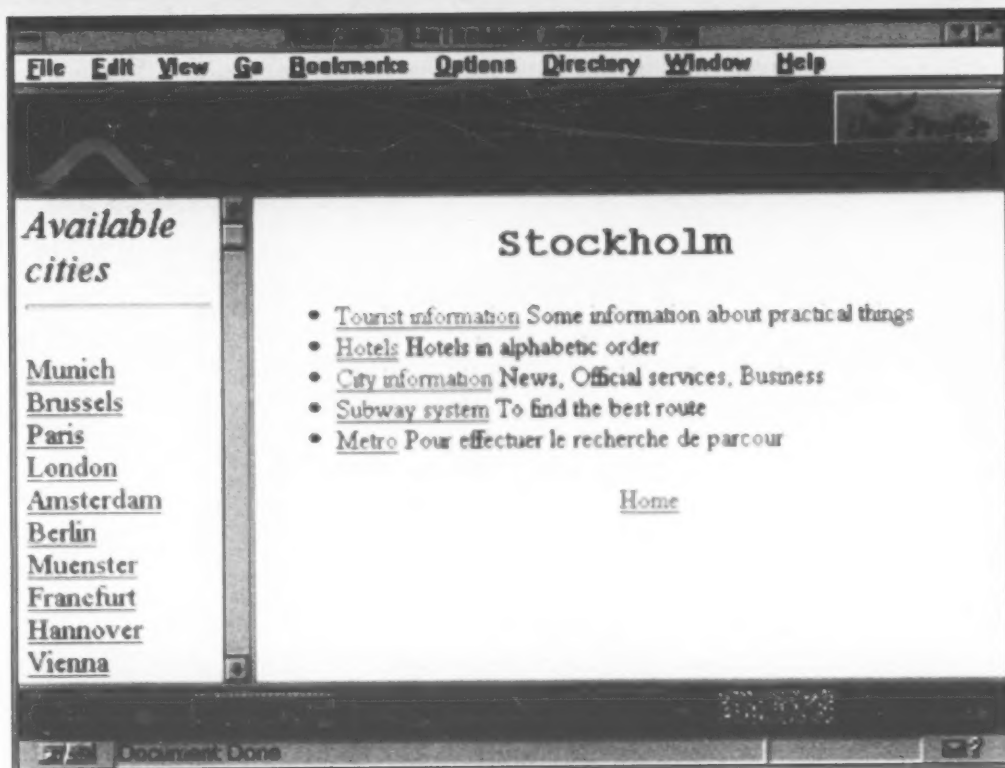


Figure 1: The Swedish application.

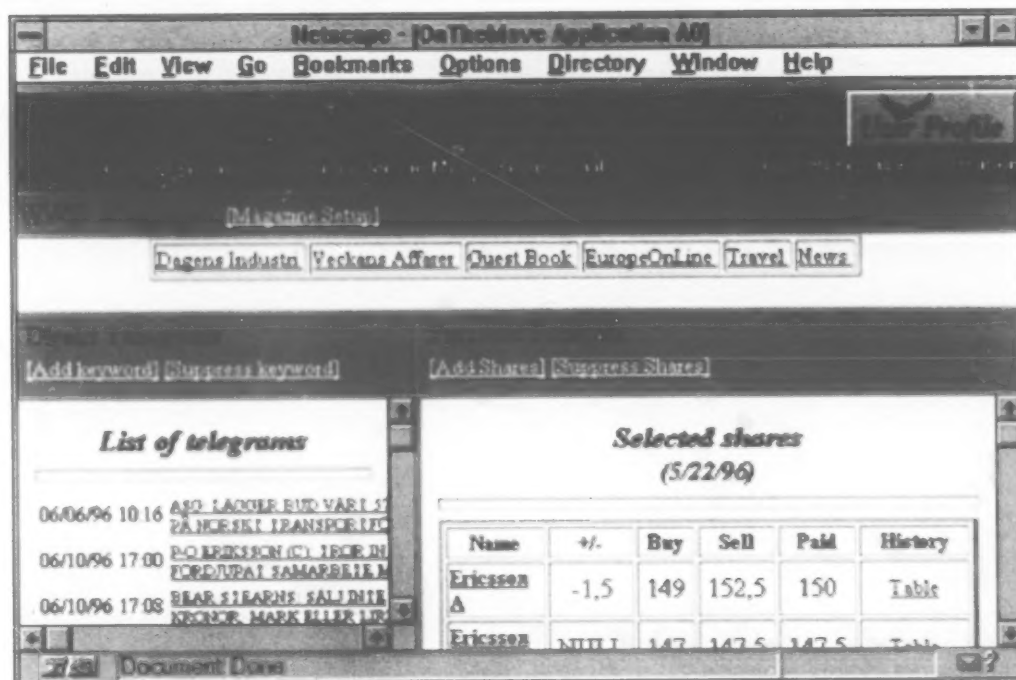


Figure 2: The German application.

The German system is based on existing databases of travel information, obtained from the Burda group, as well as news from Europe Online and links to Internet sites. The primary goal of this system is to manage the localisation of information. To do this, the server retrieves the location information from the client and automatically transmits the relevant pages.

The system will be tested by two users at the time. In the Swedish test site they use portable computers with a wireless local area network (LAN) connection to simulate UMTS capabilities (the wireless LAN having a capacity of 2 Mbps), and a mobile telephone to test GSM. In the future, the middleware will automatically manage the handover between the two. In the German test site the test uses GSM and the German Modacom network.

The test set-up has been provided to users during a field trial in June to September 1996, and a modified and developed version will be used during the first field trial of the middleware software in March 1997.

Several of the issues raised above are addressed in this system.

In the next phase of the project, these two parts of the system will be merged and create a single, multilingual, multi-national, multi-locational information system. However, using and building the system has taught us a lot about the questions posed above. We will revisit them and see how the current system addresses the different aspects of mobility, and how we can handle them even more efficiently in the future.

The system addresses the points and above as follows:

- *The network capacity is low and the user can move from an area with good reception to one with bad, or none at all, with very short notice.* The system manages this by transmitting information as text when required. It must also be able to change from graphic to textual presentation when required. This changeover is currently handled manually but in future versions will be managed through the On the Move middleware. This also applies if the user changes terminal. This, in turn, means that there must be a personal connectivity profile that changes when the users changes terminal, network bandwidth, location and other aspects of his connection.
- *The terminal itself has several limitations.* The graphics have to be clear and there have to be alternatives to the graphics. Graphics themselves must be presented in such a manner that they are readable on a black-and-white terminal, not only for users with black-and-white monitors but also for users who are colour blind (Ref 5).
- *In the case of a news service, the news feed must be kept coming even if the user moves to an area of bad reception. If the user is cut off from reception, some mechanism must be provided to handle the information.* This impacts presentation more than actual information management. We have solved it by only presenting the headlines and letting the user click to go to the actual text. In this manner the backlog of telegrams when the user has been switched off for a period of time becomes easily manageable and it is possible to skim through them. Also, as information is personalised the number of items transmitted is far lower than in an unfiltered news service.
- *If this information is automatically fed back to the sender from the receiver, the information presented can change as he or she moves from one place to another. How does the system receive coordinates?* Currently this is managed manually. In the future versions it will be handled through the middleware.
- *What are the boundaries of the current information set? What is the resolution of the database and how is this fed to the sender of information?* This is one of the thornier questions in information management. How to map objects in a database to a set of coordinates is by no means self-evident. While there are standard methods for coordinate representation, the storing of objects varies from one Geographical Information System (GIS) to another.
- *What if he is in a high-speed plane, or a train?* In this case the user is on the move (pun intended) and the coordinates will change very rapidly. Indeed, they will change faster than a GSM system can cope with, as these are only developed to handle handover from one cell to another of devices moving at speeds of less than 200 km per hour (Ref 20). How to solve this is not something we have tried to address in the current system.

4. Information of interest while on the move

Without deeper studies into information services, market surveys, focus groups and so on it is very hard to say anything definite about what kinds of information users feel are interesting while mobile. Some aspects of user behaviour have become evident, however.

4.1. Terminal mobility impact on information demand

In a system where the user accesses existing information services, be they internal or external to the company, the fact that he is mobile will put demands on the presentation of information. It will also create a need to ensure that the information which the user used to request from his desktop is available to him as he is mobile. If he is mobile across several time zones, the assistance in searches provided by an information broker or a librarian may not be available without great inconvenience to both, especially if the search requested is urgent. This implies the user will demand a simpler user interface that he himself can use without assistance.

It also implies that presentation should be kept simple while at the same time being appetising (Ref 10). This

may sound like a contradiction in terms but it is not. In the extreme case, the user may demand information over a mobile telephone without a computer attached in the form of voice-recognition. Presentation must be flexible enough to accede to this. Fortunately, developments are going on, especially in the World Wide Web community, that take this into account (Refs 19, 21).

4.2. Information mobility impact on information demand

While terminal mobility can be addressed within the framework of existing information services (albeit with a little squeezing), the mobility of information is not managed so easily. To present a localised information set, the information must be structured into objects and these must be tied to coordinates. In the simplest case, the object is a city or even a country, and the coordinates the latitude and longitude bounding this place. However, just looking at a map tells us this picture is an oversimplification. It is only in America that states and cities have straight edges. This translation, and how it should be managed is likely to provide a lot of headaches to future information system designers (Ref 11).

5. Obtaining information

The process of provisioning an information service such as the one described above is not simply a question of redressing existing information services in mobile suits. Aspects such as filtering of information on different personal criteria, localising information and differing presentation on bandwidth impact the information sources in sometimes profound ways.

5.1. Existing sources

As described above, the mobility of the user will impact the presentation of information profoundly.

One aspect is a trend that is also evident in fixed systems, albeit not as strongly. It is the trend towards dynamic presentation of information sets. As the user becomes mobile, the location becomes one more item of interest in the user's personal information profile. Even for the small number of users in the On the Move project experiments, this has meant that there will never be two users with one completely overlapping set of interests. To facilitate presentation we created fixed layouts in the World Wide Web into which the information was fed.

There were other aspects that put greater demand on the information structure, however. In the system there is a possibility to select the bandwidth of the network. In the low bandwidth case, the presentation will not contain video and graphics. Currently this means that we have to have two versions, one with graphics and one without. While this is nothing out of the ordinary to most Web designers, the next step is video that can adapt to lower transmission capacities. This will be covered in some more detail below.

But this adaptation is not limited to video only. Text may also be adapted, as well as pictures, not only by throwing undesired items away but also by filtering out information that is not necessary for the content or adding information that enhances understanding. A local resident might not need any reference on a local news item but a foreign visitor might (Ref 8). Such filters exist but are so far only experimental; although it is easy to see their application to mobile systems where bandwidth may be a scarce resource.

In a personal information system, it must be possible to apply personalisation to information. We did this in two ways in the different versions of the system. In the German version, the user can select news items from a menu where the subjects areas are pre-coded into the system. In the Swedish system the user inputs keywords and the server then dynamically retrieves only those items where the keyword occurs.

These two ways of managing information are suited to very different information systems. In the first case there must of course be a way to code the information with the relevant subject areas. These must be either input manually or extracted from the information. Currently they are input as the text is input. This means that the method is more suited to large databases, fairly static, with a slow update frequency (Ref 11).

The method of retrieving items based on keywords is fairly CPU-consuming but not unreasonably so. However, as the experiment has only had two simultaneous users so far there is no real way of knowing where the ceiling is (Ref 7).

But the real implication of personalisation is for traditional systems. To be able to provision information in personalised formats to a large number of users on the fly, your database has to be very flexible and very fast. Traditional sources of information like database hosts are not in my view likely to disappear, but they will need to adapt.

5.2. New information sources

An even more interesting set of problems occurs in video transmission, however. We have not attempted any user trials with this, conducting only small-scale experiments, but this is the area where bandwidth adaptation really comes into play. In the most extreme case, the video is deferred altogether; in the second most extreme, sound only is played. There is then a rising scale of throughput up to the full broadcast quality video — something that is very bandwidth consuming, using up all the transmission capacity to the mobile unit. In an MPEG coded film, different bandwidths can for instance easily be attained by successively decreasing the number of I-frames (Ref 14). Other, more sophisticated systems of video encoding are also appearing (e.g. FLIT) (Ref 6).

This has vast implications for the production and encoding of video information, demanding that all video

sequences be able to adapt to the capacity of the medium in this manner. Looking at the different encoding schemes, it is not simply a problem of translation: the video has to be encoded anew and this implies the source is still available.

6. Presentation aspects

In a personalised system the user will be able to choose not only content but also presentation. This implies that there will be no fixed layout. The *JavaBlitz* personal newspaper (formerly *Krakatoa Chronicle*) (Ref 4) is an example of this but it does not take multimediality into account, and neither does *Fishwrap*, a similar project at MIT (Ref 3). Still, it is fairly safe to assume that users will continue to demand more presentation capabilities rather than less. A system where only text is presented will not gain as wide acceptance, but this is one of the cultural dependencies we have discovered during the project. German users tend to state that they are more ready to receive information in text-only format than Swedes for instance.

To retrieve information dynamically based on keywords demands that the stream of information be either fairly slow or based on fairly short items (no more than 50–75 words, as far as we can see). This is not a technical limitation but has more to do with the user's capacity of reading items off the screen. While it is possible to read many items fast, the user will then need to concentrate on the screen continuously. Our system was designed for users who are accessing it fairly infrequently (4–5 times per day). This implied that too many items would generate information overload (Ref 7).

An interesting experience is also that users tend to start by a large number of keywords and then de-select them as the time passes. This implies that the user concentrates on his areas of interest rather than broadens them. We believe, although this has not been substantiated, that the users use the system for the things where it brings distinctive advantages, such as rapid updates in a narrow field, and not for general information where the regular news media are more convenient.

One of the more pertinent questions is multilinguality. If the user is able to create his or her own personal profile, why limit it to English? However good at English we non-English speakers get, there is a security in receiving information in one's own language. Even our test users, sophisticated though they are, preferred the Swedish information to the English (Ref 7). If an information set is to be presented to a larger user group, native in several languages, this implies that to be most comfortably accepted it has to be presented in their native languages. The personal profile has to contain the language preferences and act upon them. Incidentally, this holds true for both the surrounding navigation infrastructure (menus and so on) as well as for the information itself.

Of course, acquiring this information is a very different matter. In our system we simply buy it from news services. In an intranet it would have to be translated but would probably be so anyway. We can let the personalisation module of our system act on the information, since company names are the same in all languages; however, a table of 'business keywords' translating for instance interest in several languages could be constructed, and when the user chooses language preferences relevant keywords could be applied to the retrieval of information.

Cultural background is another aspect that heavily impacts presentation. The different impacts of colour are well known (Ref 5). Colour also has to be used judiciously with an eye to contrast. A very instructive exercise is to look at the page on a black-and-white screen.

When the user is mobile, we also have to take into account that not only may the bandwidth of his receiver be too low to receive multimedia information but it may also well be the case that his portable unit is something like the Nokia 9000 communicator (Ref 18), a device with a rather small screen but larger than that of a mobile telephone.

7. Conclusion

The On the Move project, while its main activity is to develop a software to manage the technical aspects of user mobility, has given rise to many questions concerning information use in a mobile environment, and some conclusions on how to adapt information to personalised systems or systems with varying connectivity bandwidths.

When presentation must adapt to different terminal capabilities and bandwidth, the information must be coded accordingly. Information may become dependent on a geographic location and change with the location of the user. The system must be able to change from graphic to textual presentation when required. This also applies if the user changes terminal. This in turn means that there must be a personal connectivity profile that changes when the user changes terminal, network bandwidth, location and other aspects of his connection.

Personalisation is another aspect of information adaptation. In a personalised system, the user will be able to choose not only content but also presentation. This implies that there will be no fixed layout.

To retrieve information dynamically based on keywords demands that the stream of information is either fairly slow or based on fairly short items (no more than 50–75 words, as far as we can see). This is not a technical limitation but has more to do with the user's capacity for reading items off the screen. While it is possible to read many items fast, the user will then need to concentrate on the screen continuously. Our system was designed for users who are accessing it fairly infrequently (4–5 times per day). This implied that too many items would generate information overload.

An interesting experience is also that users tend to start by a large number of keywords and then de-select them as the time passes. This implies that the user concentrates on his areas of interest rather than broadens them.

Cultural background is another aspect that heavily impacts presentation. The different impacts of colour are well known. Colour also has to be used judiciously, with an eye to contrast. A very instructive exercise is to look at the page on a black-and-white screen. The graphics have to be clear and there has to be alternatives to the graphics. Graphics themselves must be presented in such a manner that they are readable on a black-and-white terminal, not only for users with black-and-white monitors but also for users who are colour blind.

Another aspect of personalisation is multilinguality. However good at English we non-English speakers get, there is a security in receiving information in one's own language. If an information set is to be presented to a larger user group, native in several different languages, this implies that to be most comfortably accepted it has to be presented in their native languages. The personal profile has to contain the language preferences and act upon them. Incidentally, this holds true for both the surrounding navigation infrastructure (menus and so on) as well as for the information itself.

As the user becomes mobile, the location becomes one more item of interest in the user's personal information profile. Even for the small number of users in the On the Move project experiments, this has meant that there will never be two users with one completely overlapping set of interests. To facilitate presentation we created fixed layouts on the World Wide Web into which the information was fed.

The information will want to receive the information he is used to requesting from his desktop while he is mobile. This implies the user will demand a simpler user interface that he himself can use without assistance.

User mobility forces the information provider to adapt to several technical limitations, such as the network capacity being low and the user being able to move from an area with good reception to one with bad, or none at all, with very short notice.

If coordinate information is fed back to the sender automatically from the receiver, the information presented can change as the user moves from one place to another. How to map objects in a database to a set of coordinates is by no means self-evident. While there are standard methods for coordinate representation, the storing of objects varies from one Geographical Information System (GIS) to another. If the user is in a high-speed plane or a train the coordinates will change very rapidly, which also may impact presentation.

In the On the Move system, there is a possibility to select the bandwidth of the network. In the low bandwidth case, the presentation will not contain video and graphics. Currently this means that we have to have two versions, one with graphics and one without. While this is nothing out of the ordinary to most Web designers, the next step is: video that can adapt to lower transmission capacities. In the most extreme case, the video is deferred altogether; in the second most extreme, sound only is played. There is then a rising scale of throughput up to the full broadcast quality video: something that is very bandwidth consuming, using up all the transmission capacity to the mobile unit. In an MPEG coded film, different bandwidths can for instance easily be attained by successively decreasing the number of I-frames.

This has great implications for the production and encoding of video information, demanding that all video sequences be able to adapt to the capacity of the medium in this manner. Looking at the different encoding schemes it is not simply a problem of translation: the video has to be encoded anew and this implies the source is still available.

But this adaptation is not limited to video only. Text may also be adapted, as well as pictures, not only by throwing undesired items away but also by filtering out information that is not necessary for the content or adding information that enhances understanding. A local resident might not need any reference on a local news item but a foreign visitor might. Such filters exist, but are so far only experimental; although it is easy to see their application to mobile systems where bandwidth may be a scarce resource.

Johan Hjelm
Bonnier Media Lab
Bonnier Building
Torsgatan 21
Stockholm 113 90
Sweden

References

- [1] ACTS Mobile Domain home page, http://www.rus.uni-stuttgart.de/SONAH/Acts/mobility/home_1.htm.
- [2] Amir-Alikhani, H. et al. (unpublished) MASE Problem Space (On the Move Project Deliverable D03).
- [3] Bender, W. (1994) Read all about it in the Daily You, *Communicating Business*, Winter 1994/95.
- [4] Bharat, K. Demo of JavaBlitz at <http://www.cc.gatech.edu/gvu/people/Phd/Krishna/JavaBlitz.html>.
- [5] Brown, J.R. et al. (1995) *Visualization*, John Wiley & Sons, New York, p. 133 ff.
- [6] Danskin, J.M. (1995) et al. Fast lossy Internet image transmission, in *Multimedia 95, Proceedings*, ACM Press/Addison-Wesley, Reading, MA.
- [7] Deeke, F. et al. (unpublished) *Experiment 0 Analysed Results Issue* (On the Move Project Deliverable 18); unpublished.

Online Information 96 Proceedings

Page 290

- [8] Elo, S. (1995) *Augmenting text: good news on disasters*, in DAGS95: *Electronic Publishing and the Information Superhighway*, Boston, MA. May.
- [9] Europolitan home page: <http://www.europolitan.se>.
- [10] Harmer, J. *et al.* (unpublished) *Scenarios for mobility* (On the Move Project Deliverable D04).
- [11] Hjelm, J. *et al.* (unpublished) *Business information service definition* (On the Move Project Deliverable D05).
- [12] ITU World Telecommunication Development Report (ISBN 92-61-05661-X), pp. 2-3 (printed edition), also at <http://www.itu.ch/WTDR95>.
- [13] Julich, B. *et al.* (unpublished) *On the Move experiment 0 Infrastructure Issue* (On the Move Project Deliverable D14).
- [14] de Lameilleure, J. (1996) A scalable video coding for fast sequence previewing in multimedia browsing, in *Proceedings of ECMAST 1996*, Université Catholique de Louvain, Louvain-la-Neuve.
- [15] Mobile Communications International, May 31, p. 72. IBC Publishing, London, 1996.
- [16] Negroponte, N. (1995) *Being Digital*, Alfred A. Knopf, New York, p. 164.
- [17] Nelson, T. (1987) *Computer Lib/Dream Machines* (revised and updated edition), Microsoft Press, Seattle 1987, Dream Machine part (flipside).
- [18] Nokia WWW server: <http://www.nokia.com>.
- [19] Schmandt, C. (1992) From desktop audio to mobile access: opportunities for voice in computing, Chapter 8 in H. Rex Hartson and Deborah Hix (Eds.) *Advances in Human-Computer Interaction*, Ablex Publishing Corporation, Norwood, NJ, pp. 251-283.
- [20] UMTS & MBS System Requirements (UMTS Vision), European Commission, DG XIII RACE Mobile Project Line, at http://www.comnets.rwth-aachen.de/Conferences/C4P/race_summit_abstracts.html#UMTS.
- [21] W3 web server, especially <http://www.w3.org/pub/WWW/Style/Activity>



Online scientific and technological information in Nigeria: prospects and possibilities

R.O. Sodipe

Federal Institute of Industrial Research, Oshodi (FIRO), Lagos, Nigeria

Abstract: Nigeria is faced with developmental problems, most of which can be solved with the aid of scientific and technological information contained in books, journals and patents, largely obtainable from the advanced countries. There are academic and research institutions generating information that could enhance the economic development of the country but they are not sharing information. A few of them have computerised and they can form the nucleus of some national networks. The Federal Ministry of Science and Technology has commissioned a science and technology databank project to bring together the databases in the individual research institutes into a network. The university libraries are currently receiving assistance from the World Bank and it is hoped that part of the aid will be used to produce a networked computerised union catalogue. Telephone lines in Nigeria are inadequate qualitatively and quantitatively. Digitalisation of the lines is in progress. There are as yet no public-switched data lines with packet switching supports to make online information cheap and affordable to researchers. Dial-up access to the Internet is not yet possible and has to be done through service providers outside the country.

Keywords: Nigeria, telecommunications, computer networking, research and development, online information, Internet, science and technology, CD-ROM

1. Introduction

Nigeria is a technologically backward country perpetually encumbered with those social, and economic problems usually associated with technological backwardness. With a population of over eighty million, the country needs all the technology and technical know-how as well as show-how it can get to be able to attend to the myriad problems facing it. The needed technical knowledge must come through scientific and technical information (STI) either indigenously developed and locally available or acquired from external sources. It is the experience of Nigeria, like the other underdeveloped countries, that in spite of the R&D projects undertaken in the universities and research institutes, most of this STI comes from the developed countries through the printed media such as books, journals, patents, standards and so forth.

It will be interesting to know the number of foreign scientific and technical (S&T) journals to which Nigerian institutions used to subscribe. Unfortunately this author has not been able to acquire this knowledge and the figures available at the time of writing as contained in published National Union Catalogues are too old. Suffice to say that the universities, polytechnics and research institutes subscribed to thousands foreign journal titles in S&T. Ditto for books, patents, standards and other publications.

But that was yesterday, when there was an oil boom and Nigeria enjoyed a favourable foreign exchange reserve position. Today, with the country's foreign debt standing at an all-time high it has become impossible for all but a few of the institutions either to subscribe to foreign journals or to buy foreign books. This has been the case for many years now and it is not uncommon to find foreign journals published seven or more years ago still being displayed as the current issues in some Nigerian libraries. However, some libraries are in a privileged position and are able to be up-to-date in their collections. It is therefore reasonable to advocate resource sharing as the only way of solving Nigeria's problem of scarcity of literature in the area of S&T.

2. Sources of STI in Nigeria

STI information is indisputably one of the prerequisites that Nigeria needs in its struggles to become recognised internationally as a developed country. In recognition and admission of this crucial fact universities of technology, polytechnics and research institutes have been established to provide information that can be used to solve the country's problems.

As of now there are in the country 36 universities, 20 polytechnics and 24 research institutes, where research and development work is done and technological breakthroughs and inventions are made.

3. Physical and environmental handicaps

These institutions are separated from each other by long distances, cut off from each other by bad roads, poor postal services, bad telecommunication lines and poor transportation services. Between them there is a wide information gap created by these infrastructural handicaps which makes operation of traditional library cooperation schemes difficult and almost beyond human endurance. Consequently, these institutes are not exchanging data or sharing resources among themselves. Yet a means must be found through which the few internally available, up-to-date information materials possessed by the privileged few can be jointly used by all who are interested, as and when desired. The means is to be found in computerisation and networking.

4. Computers to the rescue

In spite of the generally poor state of physical, environmental and infrastructural facilities, some institutions are preparing to go into computerisation, some are in the initial stage of computerisation and some have fully computerised their operations. Among them are the International Institute for Tropical Agriculture (IITA), Nigerian Institute of International Affairs (NIIA), Federal Institute of Industrial Research, Oshodi (FIIRO), National Office for Technology Acquisition and Promotion (NOTAP), Raw Materials Research Development Council (RMRDC), National Institute for Pharmaceutical Research and Development (NIPRD) and National Research Institute for Chemical Technology (NRICT).

The university libraries which for many years were unable to purchase books due to scarcity of funds have been enjoying some assistance from the World Bank under which their book requirements are being satisfied. The cataloguing of these books in the individual libraries is done with the computer using a common software called TINLIB. With this, a computerised union catalogue of university libraries' book holdings should emerge and the university libraries will be able to gain online access to each others' catalogues, if and when they are networked.

5. Network(s) development

The essence of linking up computers in networks is to enhance data exchange and information resource sharing among corporate bodies and individuals, nationally and internationally. In order to gain maximum benefit from such cooperation it behoves all participating entities to contribute to the pool. This means that any participant in the arrangement is expected to possess its own local resources. The proverb 'charity begins at home' is very apt here. Nigeria therefore is obliged to build its own national networks of computers in which will be resident the databases of scientific/technical, statistical, economic and educational information which have been produced locally by the appropriate institutions.

In the area of STI very little is being done towards networking of computers, although some of the universities and research institutes have been producing local databases. At FIIRO, different databases have been produced relating to the following:

- on-going R&D projects;
- production raw materials inputs;
- production process techniques and technologies;
- machinery and equipment.

At RMRDC, databases have been produced on Nigerian raw materials availability and processing. NOTAP has databases on patents and patented technologies. These three organisations have some common grounds of activities; they were within each other's reach in Lagos before recently RMRDC moved to Abuja. But ironically, the three are informationally as far apart as they are geographically near. It is heartening however to note that the Honourable Minister of Science and Technology has observed this and official action is being taken to rectify the situation.

6. Some of the existing network types

IITA provides the best case of computer networking and online communication among institutions of its type in Nigeria. Workstations are linked into LANs which are further linked to form a CAN (Campus Area Network) using fibre-optic cables. A few standalone workstations outside the campus are linked to the network with radio waves using a technique in which an external antenna is pointed in the direction of the server to be connected. For its international communication IITA uses Internet e-mail. The software used is Novell network version 3.12 and it is the software that builds up the servers.

NACETEM (National Centre for Technology Management) in Obafemi Awolowo University, Ile-Ife uses coaxial cable for its network within the university.

ICTP (International Centre for Theoretical Physics) is executing a LAN project at the same place using radio waves within the university. It is planned to link up the universities in the country and to link up with the Internet via a gateway in Italy.

FMST (Federal Ministry of Science and Technology) has embarked on a project to network all of the research institutes under it and to connect them to a server to be located in the office of the Honourable Minister in Abuja, the new Federal Capital.

7. CD-ROM use in Nigeria

Another computer device that is being used in some Nigerian institutions to make good the shortcomings in information supply is the CD-ROM. Due to lack of funds to subscribe to the databases in host computers in Europe, USA and other countries, and due also to the lack of the telecommunications facilities required for effective link-up with these databases, CD-ROMs are being resorted to as a compromise solution to, and as an alternative method of, gaining access to foreign databases.

The CD-ROM stocks of individual libraries are appropriately geared towards the subject specialisation of their respective parent institutions. Thus we have at NOTAP a CD-ROM on European and world patents. At IITA there are more than 25 CD-ROM databases covering agriculture and related sciences. FIRO can boast a number of CD-ROMs dealing with science and technology generally and process technologies specifically.

The importance of CD-ROMs in Nigeria stems from two considerations. Firstly they have brought the otherwise inaccessible foreign databases (some of them) to the doorsteps of Nigeria. Secondly they have helped to actualise the idea of information resource sharing envisaged when some institutions embarked on computerisation, because by their compactness they lend themselves to interlending more than diskettes.

Information retrieval from these CD-ROM databases is done in the batch mode in all the libraries, as networking environments for online information retrieval from them are yet to be created. IITA is in the forefront in information computerisation in Nigeria and it plans to have CD-ROM in a network environment. This institute is therefore expected to do online information retrieval from CD-ROM in the near future.

8. Impediments to networking and online information in Nigeria

The biggest factor militating against networking and online information in Nigeria is the telephone line. Its supply is quantitatively inadequate as only a small percentage (below 5%) of the population has access to the telephone (Ref 1). Its transmission quality is generally poor as most of the lines are still analogue. Rental and call charges are prohibitively high and beyond what most middle class people in research and academics can afford.

8.1. Other problems

Computerisation together with its other technical dimensions is an expensive undertaking requiring huge sums of money to begin and to keep going. With most of its hardware and software produced outside Nigeria, payment for its cost inevitably has to be made in foreign exchange. This poses a big problem for successive governments and it will take an administration with deep faith in the value of information to have the *political will* to spend the much money that is required.

Funds scarcity is another problem. Nigeria is a debtor country with its foreign reserve deficit currently running at an all-time high. Under these circumstances it is difficult for the government to commit the country's lean foreign exchange earnings to make the Internet accessible to Nigerians, especially if the political will and the conviction is lacking as it seems to be.

9. Online telecommunication facilities available in Nigeria

The official provider of telecommunication facilities in Nigeria is NITEL (Nigerian Telecommunications PLC). It is being gradually converted into a private company and in pursuance of this some of its services are being transferred to commercial companies known as service providers. The following are the NITEL services available through service providers:

- *Leased line.* Voice or data lines can be leased. The cost depends on the baud rating but it is known to be high and beyond what the academic and research institutions can afford. It is therefore the exclusive preserve of commercial houses, especially the multi-national companies. It operates on a point-to-point basis and can be used for both local and international online access. The analogue leased line can presently support a maximum speed of 28.8 kb/s. Its digital counterpart can support various speeds varying from 64 kb/s, 2 Mb/s, 8 Mb/s and 34 Mb/s;
- *Public telephone line.* This is the least attractive on account of the poor quality transmission caused by

noise and distortion. In order to use it for extra-territorial access it needs to be fitted with IDD (International Direct Dialling) support;

- *Microwave link.* Online information can be achieved through radio transmissions. Instead of the telephone serving as the medium of communication, high quality radio microwaves can be used. In order to connect long distances in a wide area network, repeaters or signal boosters are installed as a pass-it-on device from origin to destination;
- *VSAT (Very Small Aperture Terminal).* This is a satellite service;
- *Electronic mail.* This is offered through the Internet and is being increasingly used by universities and research institutes;
- *Packet switching.* NITEL has just announced the availability of a packet switching service. This service is for leased line users to enable them to access multi-points instead of the point-to-point. As more and more telephone lines are digitised, packet switching will be brought to the public telephone domain and Nigeria will then have public switched data lines (PSDL).

It is apparent from the above that public switched data lines (PSDL) are not yet provided in Nigeria. The existing public telephone lines do not have packet switching support to make possible multiple use and cost sharing among users, as happens in the advanced countries. As a result the cost of online communication is still prohibitively high and beyond what scientists and researchers or even their institutions can afford.

10. Internet

It is not yet possible for Nigerians to gain dial-up access to the Internet since there is no node or gateway in Nigeria presently. The practice now is for those desiring access to go via a node in any of the advanced countries. Efforts are however being made to get Nigeria to hook onto the Internet in the not too distant future. In fact there is a movement towards this, spearheaded by a group known as the Nigerian Internet Group (NIG). The Nigerian government has commissioned a survey to find out the desirability or otherwise of putting Nigeria on the Internet.

11. Conclusion

Online transmission of information is still in its infancy in Nigeria and not much has been done to warrant international attention. Its use for business in the private sector is restricted as only the banks, airlines, oil companies and other multi-nationals can afford the exorbitant costs of leased lines. Its use in the universities and research institutes is prohibited by lack of funds. Of the Nigerian institutions that have computerised and networked, none has been fully funded locally and all invariably were established with foreign assistance, usually from international and inter-governmental organisations.

The problem of financing computerisation and networking in Nigerian public institutions and the problem of providing the required infrastructural facilities such as public switched data lines, Internet gateways and nodes to actualise online information are attitudinal problems, born of poor understanding of the importance of information and resulting in the low priority accorded to it in our order of things. It takes a government with vision and with the right attitude to information to build the kind of electronic infrastructural backbone like the Internet, which now supports online information in the advanced countries, for colossal sums of money. It is not for lack of money therefore but rather for lack of interest and understanding that Nigeria lags behind in this laudable global competition on the information superhighway, where knowledge and experiences of people scattered all over the world are brought together for mutual sharing.

The universities and the research institutes on their part do not appear to be doing enough in documenting their outputs and disseminating them in the modern media for local and wide area consumption. Their libraries and information units need to rise to the challenges of modern methods of information processing by eradicating computer illiteracy among their professional staff, and additionally recruiting computer experts in programming/systems analysis and data processing. Information scientists to do document scanning and analysis, indexing, abstracting and keywording are also lacking in Nigerian universities and research institutes. Without the coming together of telecommunications and computer technologies there would have been no information technology. It can therefore be said prophetically that without massive introduction and employment of the above experts in universities and research institutes, there will be nothing on which to base networking and online information.

R. O. Sodipe
Federal Institute of Industrial Research, Oshodi
P.M.B. 21023
Ikeja
Lagos State
Nigeria
Tel: +234 (0)1 521010, 522760, 523260
Fax: +234 (0)1 4525880

Acknowledgements

In preparing this paper I have used information gathered from Mrs M. Atalobhor of IITA, Mrs N. Denloye of Telnet (Nig.) Ltd and Professor S. A. Sanni of NACETEM. I am grateful to them.

References

- [1] Agbinya, J.I. (1995) Internet access, ethical and legal issues: the Nigerian Context, in *Proceedings of a Two-Day National Workshop on Internet Connectivity*, Lagos, Nigeria, 6- 7 September 1995.
- [2] *Proceedings of a Two-Day National Workshop on Online Services: The Place of Nigeria in Global Information Technology*, Lagos, 6-7 May 1996, Simulation Systems Ltd.



The 'superhighway': a revolutionary means of supporting collaborative work

C.A. Grant and T.M. Scott

University of Ulster, UK

Abstract: *This paper looks at how Advanced Communication Technologies are revolutionising the way collaborative teams can work together. It represents great opportunities for collaboration in the fields of research, commerce and education. In particular we will discuss how electronic collaboration is becoming an essential component of business and university life, and how Internet technology is being recognised as a powerful new means to improve the productivity of working groups.*

High-speed communication networks are a powerful tool for promoting successful international partnerships. Using high-speed bandwidth to communicate, collaborative research and development work can be progressed efficiently (without waiting until the next scheduled meeting). This paper illustrates how the authors have gained advantage by exploiting current technology to form successful partnerships in this new, high-speed, 'no boundaries' information age.

Keywords: Internet, high-speed communications, ISDN, European R&D, electronic collaboration.

1. Introduction

Promoting collaboration across national, European and international boundaries is a challenging task. A variety of obstacles including physical distance, national and regional differences in language, business cultures and legal provisions continue to frustrate transnational developments. The key to a successful collaborative project is interactive communication, and the Internet is fast becoming an indispensable international communications tool (Ref 1). In cooperative projects partners often come from different environments, and possibly different regions and countries. Barriers such as geographical remoteness, lack of time in the working day, the cost of travel and differences in culture can seriously hinder regular communication and development work. The exploitation of advanced communication technologies can remove many of these barriers, and the examples of projects made possible through Internet communication are astounding (Ref 2). Internet technology is enabling people in different countries to exchange ideas, develop proposals, advance projects and resolve differences or misunderstandings without ever meeting physically.

Advanced communication tools such as electronic mail, archiving, teleworking, desktop conferencing and interactive video conferencing are becoming cheaper and easier to use (Ref 3). Web technology provides an inexpensive and easy-to-use user interface that is fast becoming a prevalent tool on most desktops. Geographically distant project teams can now plan their programmes of work by remote video conferencing, in which face-to-face discussions can be enhanced by sharing of information held in partners' respective computers. High quality pictures, sound and images can be transmitted over high-speed communication networks to colleagues the other side of the country or the other side of the globe (Ref 4). In effect, broadband communication networks can serve to facilitate communication and collaborative working.

2. The Internet

The 'information superhighway' has been the subject of a great deal of media hype and fanciful speculation. It conjures up a vision of speedy electronic networks, connecting homes and businesses to a boundless trove of information, communication, education and fun. It is predicted to have profound effects on society, the way we communicate with each other and the way we will do business. Trouble is, it isn't here yet, and chances are it will not be a reality for a while.

In the meantime, a forerunner of the information superhighway already exists: it is called the Internet and it is the world's biggest collection of interconnected computer networks. Some people argue that the Internet already is the information superhighway: it simply looks different from the television-based version promoted by the cable and telephone companies. Others say that some important features are missing: the real superhighway will

combine not just computer networks but all sorts of other communication links, from telephone and radio to satellite (Ref 5). It will be fast where the Internet is slow. Yet the Internet already offers, although in embryonic form, most of the services and technologies that cable and television companies are still a decade from delivering. You can make a telephone call on the Internet; watch a video; listen to an audio broadcast; shop; learn; and of course, communicate. It is a major tool for sharing computer resources, research issues and project ideas, and for communicating with groups and individuals.

The Internet has grown exponentially in recent years and its growth is estimated to be doubling in size and volume on an annual basis, wrapping the world in a massive computer web. At present it connects 30 million users across 130 countries and it is estimated there will be some 100 million users by 1999 (Ref 6). No communications medium or consumer electronics technology has ever grown as quickly: not the fax machine and not even the PC (Ref 5).

3. Electronic pot holes

The Internet is easy to ridicule: it can be chaotic, frustrating and intimidating, and 90% of what is available is rubbish (Ref 5). One of the biggest difficulties that Internet collaborators face is the inherent impersonal nature of the medium. Even with the aid of audio and video conferencing, it is difficult to capture the experience of a face-to-face meeting with a group of colleagues. However it is here to stay, and it will allow people to exercise one of their most basic desires: to communicate (Ref 2). They are able to do this on an unprecedented scale: globally, openly, to one person or many.

4. Superhighway infrastructure

At present the Internet suffers from data speeds that are too slow to move today's multimedia information: 'we have yet to overcome the lack of suitable, high performance, accessible communications infrastructure' (Ref 7). The problem lies with what is commonly referred to as the last mile — the last link between large scale fibre-optic networks (which can be strung along existing rights of way) and customer's homes or businesses. At present this last mile is being bridged by modem technology, which converts digital signals into an analogue format compatible with the traditional telephone (Ref 8). Although modems are widely available at a low cost and are easy to use, they are not able to cope with the increasing demand for bandwidth-hungry applications such as desktop conferencing.

A new infrastructure is underway: within the next decade users will have direct, high-speed connections to the Internet with more effective data transfer rates (Ref 9). A direct connection is one that does not involve ordinary analogue telephone lines with their relatively slow data transfer rates of 1.2 to 28.8 kilobits per second (Ref 6). As we speak a countless number of computers and computer technologies are coming together (linked by high capacity coaxial cable, fibre optics and wireless transmission systems) to create a vast information infrastructure for faster, more efficient information exchange (Ref 8).

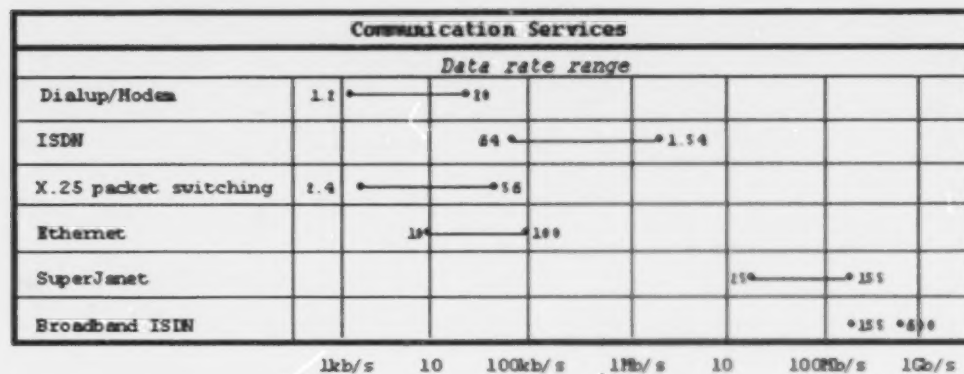
The digitalisation of telephone networks has spawned a new service, ISDN (Integrated Services Digital Network), which allows telephony, e-mail, fax, Internet/Web and full-motion video down a single connection: you get a transmission speed of 64kb per second, roughly five times faster than a standard telephone line with a modem connection (Ref 10). Even so, ISDN has been slow to take off, mainly because for a long time it was ridiculously costly and not widely available. It is only in the last couple of years (in direct correlation with the popularity of the Internet) that individuals have been turning to ISDN as an affordable access vehicle that performs nearly on a par with the expensive dedicated connections of larger organisations (Ref 10).

In competition with the telecommunications industry (ISDN providers) is the TV cable industry, which is constantly digging up the roads and pavements to lay down fibre optic cables. Fibre optic cables carry orders of magnitude more information than their copper base equivalent and the costs of the two are more or less the same (Ref 10). Hence in the not too distant future, both the telecommunications and cable networks will be capable of transporting and processing voice traffic, text and images instantly to and between all households and businesses around the world (Ref 4). This more advanced communications infrastructure has the potential to revolutionise manufacturing, research, education, health and commerce (Ref 11). It will be the medium by which companies will buy, trade, make contacts, exchange data and information, hold conferences, discuss designs and find components (Ref 11).

In the UK this main infrastructure is already in place (for academia) in the form of the SuperJANET broadband network (arguably the world's first superhighway) (Ref 12). SuperJANET is referred to as a WAN or wide area network, since a large geographical area is covered. When talking about site networks we refer to LANs (local area networks), i.e. universities are interconnected to their own departments and campuses through a local area network such as Ethernet. The universally used Ethernet is a high-speed network which currently works at a data transfer rate of 10 Mb/s (see Table 1). Ethernet is compatible with SuperJANET and forms part of the overall network infrastructure. As the name suggests, SuperJANET — the Super Joint Academic Network — is an advanced high-speed optical fibre network linking over 60 universities and research laboratories in the UK (Ref 13). SuperJANET optimally works at a speed of 155 megabits per second. To put this capacity in context, this means that the equivalent of a 5500 page report can be sent in less than one second (Ref 13). High bandwidth networking at these speeds allows the remote use of advanced multimedia technologies hitherto only available

to users on a local basis. SuperJANET is the UK's portion of the global Internet and ultimately it will become part of a global high bandwidth network — an information superhighway which knows no national boundaries (Ref 4).

Table 1: Range of communication services.



5. Economics of the Internet

A simple question that a lot of people ask is: if the Internet runs over telephone lines, why does it cost the same to send an e-mail message around the world as it does to send it next door? The first reason why long distant messages costs no more than local ones is that the Internet, although it runs on telephone lines, uses them much more economically than voice calls do. A voice call is an analogue signal which needs lots of electronic space to avoid interference: it takes up an entire line for the duration of the call. By contrast the Internet is digital, so data bits — ones and zeros — can be compressed (Ref 14). Secondly, Internet data is split up into packets, which do not need a line to themselves. Packets from hundreds of sources are mixed up by the computer and shoved down the telephone line in a jumble. The router at the other end of the line receives each one, reads its address and sends it in the right direction (Ref 14). When you make a telephone call, you are consuming a scarce and expensive resource: a whole line. In comparison when you sending a message on the Internet, you are sharing a plentiful and cheap resource: the entire bandwidth on the line (packets of information are just a drop in a passing river). There is a third reason: telecom pricing is a notorious scam (Ref 14). A large part of the price of a telephone call (often more than 40%) goes to the recipient's telephone company for taking it the last few miles. Through a complicated accounting scheme known as 'settlements', telecom companies exchange billions of pounds each year for the local component of international calls. The Internet bypasses all this: it usually operates on leased lines, out of reach of national telecoms accounting which would find it hard to track and cost a constant stream of ones and zeros.

6. Collaborative working and Internet services

In the case study which will be described, you will be made aware of the various services and applications available on the Internet which enable collaborative working to be carried out. Some of these services are explained here:

- **E-mail.** Electronic mail is by far the most widely used and most basic service available on the Internet. It allows you to send messages to colleagues all the world in a matter of minutes: it is convenient, inexpensive, reliable and fast (Ref 15). In a company or university, an example of everyday use of e-mail is to announce conferences, schedule meetings, arrange appointments between employees and managers, and distribute committee meetings. E-mail is a conduit for sharing hypotheses and results, preparing draft documents and for discussing project issues. The uses of e-mail are endless: in fact it has become an essential tool in most collaborative work and certainly in everyday academic life.
- **File Transfer Protocol (FTP).** FTP is a useful application that allows you to connect to a host computer and copy some of the host's files to you own computer. FTP provides access to hundreds of gigabytes of programs, text, complex graphics and other miscellany (Ref 16). In collaborative projects this is a useful

resource where a number of partners in different countries are working on different sections of joint business proposals/reports: they can use FTP to pull the final document together.

- **World Wide Web.** The appeal and success of the ground breaking World Wide Web information service is the driving force behind the explosion in populist use of the Internet. It is by far the best way to retrieve information from the Net and it is already revolutionising the way in which information is accessed and presented. It provides a very user-friendly graphical interface (and hypertext links) to connect you to information on any computer around the world. It supports numerous online services which support collaborative working (see case study).
- **Video conferencing.** Video conferencing is more an application than a service which is supported by the Internet. Video conferencing and desktop conferencing are enabling collaborative groups from different organisations and different countries to have face-to-face meetings without the expense of time and travel. In essence, advanced optical fibre networks help us achieve a rapid and effective meeting of minds without first having to secure a physical meeting of people (Ref 13).

7. Case study: electronic collaboration

The University of Ulster has established a wide range of strategic academic and industrial alliances around the world. This case study describes how research teams at the university are using the Internet to enable these valuable partnerships to be formed.

7.1. Background to European R&D Partnerships

In Europe the Fourth Framework programme provides a major incentive for universities and industry to work together. It supports industrially relevant collaborative research projects by providing direct access to 50% funding. In order to secure funding for European collaborative projects, one has to find partners and subsequently submit a detailed project proposal before the R&D programme deadline (Ref 17). Preparing a proposal involves a substantial investment of time and money, and it is expensive to coordinate the work of a widely spread project team (especially considering the fact that on average only one out of every seven proposals are approved for funding) (Ref 17). Below we will describe a case history which reflects real experiences from a collaborative European project:

7.2. Case study

7.2.1. Step 1: the idea

A small manufacturing company in Sweden needed to implement a shop floor production control system but did not have the expertise. It approached a Research Institute in France with this strategic problem. It was decided to form a consortium and submit a project proposal to the Fourth Framework programme. In order to qualify for EU funding, it is advantageous to have a balanced consortium (of both industrial and academic partners) from about four different EU member states.

7.2.2. Step 2: finding partners

The French partner carried out a partner search using the CORDIS online information service, which is available on the Internet and accessible through the World Wide Web. CORDIS provides a centralised source of information for disseminating, assisting and promoting cooperation on EU R&D projects. It contains details of over 24,000 organisations either seeking or offering themselves as partners for collaborative projects (Ref 18). Fortunately, one of the main research groups at the University of Ulster had submitted a profile of expertise electronically about four months previous. The French and Swedish partners were looking for expertise in scheduling technologies: at this point the University of Ulster was contacted by electronic mail, interest was generated and commitment gained.

7.2.3. Step 3: R&D programme

Further ideas for the project were exchanged using electronic mail. The next step was to find out which Fourth Framework sub-domains the project idea fitted into (i.e. which sub-programme was most likely to fund the project). Again, from a desktop, one can tap into the information database in Brussels and look up information on the various sub-programmes of the Fourth Framework. The appropriate sub-programme was called 'Integration in Manufacturing'. An electronic copy of the programme call and the information package with application form were downloaded from the Internet and printed out. A potential problem arose: there was only two and a half weeks left before the submission deadline. This is when the technology really went into full drive — we had to get at least three more suitable partners on board, write up the project proposal and get it to Brussels before the deadline.

7.2.4. Step 4: networking

'Sweeping across the landscape like a firestorm, the Internet has revolutionised networking' (Ref 19). The University of Ulster group accessed the CORDIS partner search database to look for an industrial end-user partner in either Northern Ireland or the Republic of Ireland. The database came up with several suitable manufacturing companies, which were all contacted and another small electronics manufacturer responded. The French partner used CORDIS again to look for another suitable partner, this time with expertise in database technologies. A French software house joined the team. The consortium was complete. With five partners on board, it was necessary to set up a meeting. We needed to document project objectives, allocate work packages, estimate project costs and human resources, and so on. For one reason or another, two partners could not make the proposed meeting in France: an alternative solution (which saved time and travel) was to have two video conference sessions between France, Sweden and Northern Ireland. The small manufacturing companies from Sweden and Ireland had access to video conference facilities through nearby universities. An agenda and various documents were exchanged electronically (via FTP) before going online. The conference sessions permitted partners to meet face to face, and plan out the project and share ideas, skills and resources over the network rather than in person.

7.2.5. Step 5: Working on proposal

Substantial work on the project proposal was completed in each partner's own country. The Swedish partners became the project leaders and coordinated the compilation of the final proposal. Partners availed of services offered by the Internet to look up state of the art facts. An Internet search was also carried out to verify that the project had not been done before. Using computer-based language translators and interpreters helped common understanding between partners. The French and Swedish partners wanted to set up an electronic brainstorming session. The University of Ulster had used this groupware application on a number of occasions: it enables distributed synchronous information sharing (i.e. activities taking place at the same time in different locations (Ref 20)). It helped us achieve significant productivity, especially during the task of costing the project (where opinions and partner costs varied widely): the brainstorming sessions (using the groupware application) let partners enter comments and ideas simultaneously and anonymously. Within a week, various sections of the project proposal were transferred byte by byte over the Internet to various partners for the purpose of agreeing its content, making changes, and checking it was complete and correct before the final proposal was sent to the European Commission. Once submitted, partners continued to communicate on a regular basis using e-mail. Three months later the project leader received notification that the project proposal had been successful and project work has since commenced.

8. Benefits of linking up

Analysing this case history, it would have been impossible to complete the project proposal in time if the partners had not been connected through the global Internet. Partners used e-mail extensively: it was cheap, fast and reliable. At different stages of the proposal writing, FTP was used to transfer the draft proposal from partner to partner: the fact that the document was in electronic form made it easy to edit (imagine 20 years ago having to fax the draft document or send it via snail mail, i.e. regular post). The electronic brainstorming session was crucial in solving decisions regarding estimating costs of the proposed project.

Communicating using advanced technologies significantly reduced the costs of collaborative working. It is estimated that five years ago a transatlantic video conference call (ISDN, single line) cost £800 an hour: now that same call is less than £40 per hour with substantially better quality (Ref 3). Video conferencing improved productivity by reducing time lost to travel and efficiency lost to jet lag. In the case history above, time was the crucial element: it was not possible to get all the partners together at a meeting for the proposed meeting date, but it was possible to get them to meet electronically to implement important project decisions.

9. Conclusion

Electronic technology and high-speed communications are no longer simply aids to collaborative working. In this information era, information that is speedy, accurate and in a form that enables people to make decisions and work productively together across geographical distances is the key to cost-effective efficient collaboration. The future will offer ever increasing capabilities for electronically enhanced collaborative working. 'Throughout the world, information and communication technologies are generating a new industrial revolution already as significant and far reaching as those of the past . . . This revolution adds huge new capabilities to human intelligence and constitutes a resource which changes the way we work together and live together . . .' (Bangemann Report, Brussels 1994). Failure to meet the challenge of this rapidly expanding information environment will be reflected in failure to survive in the Information economy of the future.

Caroline Grant
Faculty of Informatics
University of Ulster
Co. Londonderry BT48 7JL
Northern Ireland
UK
Tel: +44 (0)1504 375236
Fax: +44 (0)1504 375251
E-mail c.grant@ulst.ac.uk.

References

- [1] SPRINT. (1995) Communication in the network, *Good Practice in Managing Transnational Technology Transfer Networks*, **1**, 71-75.
- [2] Cranor, L.F. (1995) Internet collaboration: good, bad, and downright ugly, *The Economist*, July issue.
- [3] Snell, M. (1994) Picture this: Videoconferencing, *Computer*, **21**, 10-11.
- [4] Hartley, D. (1995) Whither the information superhighway, *Computer Bulletin*, **7**(2), 2.
- [5] Anderson, C. (1995) The accidental superhighway: a survey of the Internet, *The Economist*, July issue, <http://www.economist.com/surveys/>.
- [6] Meltsner, K. (1995) Understanding the Internet: a guide for materials scientists and engineer, *JOM*, **47**(4), 9-10.
- [7] Maule, R. (1994). Smart valley infrastructure, *Internet Research*, **4**(2), 48-50.
- [8] Jaret, P. (1995) Grabbing a lane on the information highway, *EPRI Journal*, **20**(1), 21-23.
- [9] Meltsner, K. (1995) Understanding the Internet, *Journal of Mineral Metals and Material Society*, **47**, 9-13.
- [10] Frankel, D. (1995) ISDN reaches the market, *IEEE Spectrum*, June, 20-25.
- [11] Barthrope, F. (1994) Hitch a ride on the superhighway, *Professional Engineering*, **7**(13), 12-13.
- [12] Barden, L. (1995) The university, the superhighway and regional economic development, *Industry & Higher Education*, **9**(5), 315.
- [13] UKERNA (1996) What is SuperJANET, <http://www.ukerna.ac.uk>.
- [14] Anderson, C. (1995) Freeloading as a way of life: a survey of the Internet, *The Economist*, July issue, <http://www.economist.com/surveys>.
- [15] Zelingher, J. (1995) Exploring the Internet, *M.D. Computing*, **12**(2).
- [16] Digiorio and Richert (1994) E-mail, the Internet and information access, *Seminars in Diagnostic Pathology*, **11**, 294-300.
- [17] Commission of the European Communities (1994) *EC RTD: Insight into handling of Project Proposals*, pp. 5-9, ISBN 92-826-7076-7.
- [18] Commission of the European Communities (1996) CORDIS: online, *Innovation & Technology Transfer*, **1**, 12-13.
- [19] Cavallini, A. (1996). Community nets: the next generation, *Computer*, **29**(2), 92.
- [20] Palmer *et al.* (1994), Computer-supported co-operative working, *Computer*, **21**, 16.



The telecommunications stranglehold on Europe's information use: practical constraints for SMEs and economic assessment based on cost models

Jean-Claude Delcroix

DECADE sa, Belgium

Abstract: *This paper starts from research we did at DECADE on the requirements of smaller organisations on the one hand and on telecommunications costs on the other. In the new information age, competition is forcing all companies and increasingly also SMEs to get and handle a lot of information. Speed and costs are two key competitive factors as SMEs operate with rather short term constraints and have very low purchasing budgets. Traditional online services appear far too expensive to use and very difficult to learn. The World Wide Web matches the knowledge acquisition paradigm of end-users and SMEs much better than traditional online services. From this perspective, in most European countries the current offer of online information services and related telecommunications methods is seriously lagging behind compared to the US. In the paper we compare the network costs and the Internet-based services in Europe and in the US. We also provide some data about Japan. We show graphics with price comparison for leased lines and for Internet access. We investigate what are the current shortcomings: is it the distance that matters (cost per kilometre or mile)? Is it the cost of bandwidth? We also have a look at the likely evolution towards TCP/IP data networks and bandwidth costs per user. In our conclusion we address key economic issues regarding information publishers, Internet/intranet service providers and of course what European users might do.*

Keywords: telecommunications costs, SME requirements, online services, Internet, Europe, USA

1. Introduction

There is a general feeling that European telecommunications are delaying the introduction of new information services. Is that right? How do European costs and bandwidth supply compare with the US? What is the impact of the current status on European firms and SMEs in particular? Are bandwidth constraints hampering multimedia technology? These are some of the many questions that are currently arising about online information. This paper tries to provide some answers and to show the way ahead in our fast changing world.

Our views result from research work at DECADE on the requirements of smaller organisations. We have investigated how they search information and what are affordable costs. We complemented these findings by making a cost comparison for telecommunications services and Internet access, using a model.

2. Information searching by end-users and SMEs

In the new information age, competition is forcing all companies and increasingly also SMEs to get and handle a lot of information. Speed and costs are two key competitive factors as SME operate with rather short term constraints and have very low information purchasing budgets. Traditionally end-users and SMEs do not use online information services. They search information mainly by having personal contacts and by reading articles. They are used to 'browse' through manuals and periodicals without always knowing exactly what they want to find: it is a discovery process. When they know what to search, they have difficulties in expressing their needs in online terms.

Most of the SME's information budget is made of people's time. A medium-sized company may spend as much as 125,000 ECU a year in human time for information gathering versus 12,000 ECU in information sources, of which 5000 ECU is for trade associations fees.

Traditional online services appear far too expensive to use and very difficult to learn. The less the user under-

stands the command language and the database content the higher the resulting cost. Downloading large amounts of documents to browse through may cost several thousands of dollars per search. Also, it is not realistic for a company with, say, 40 office workers to have one person that can be trained to use a variety of online databases every week. Therefore — it is no surprise — the World Wide Web appears to match the knowledge acquisition paradigm of end-users and SMEs much better: that is, how they get information from other people, how they 'browse' the world of the unknown in search of ideas or products or whatever. With the Internet they get full text, images and the full world with the same protocol and user interface.

Unfortunately online information services are slow to adapt. There are many obstacles between SME requirements and online services which can be summarised in Figure 1.

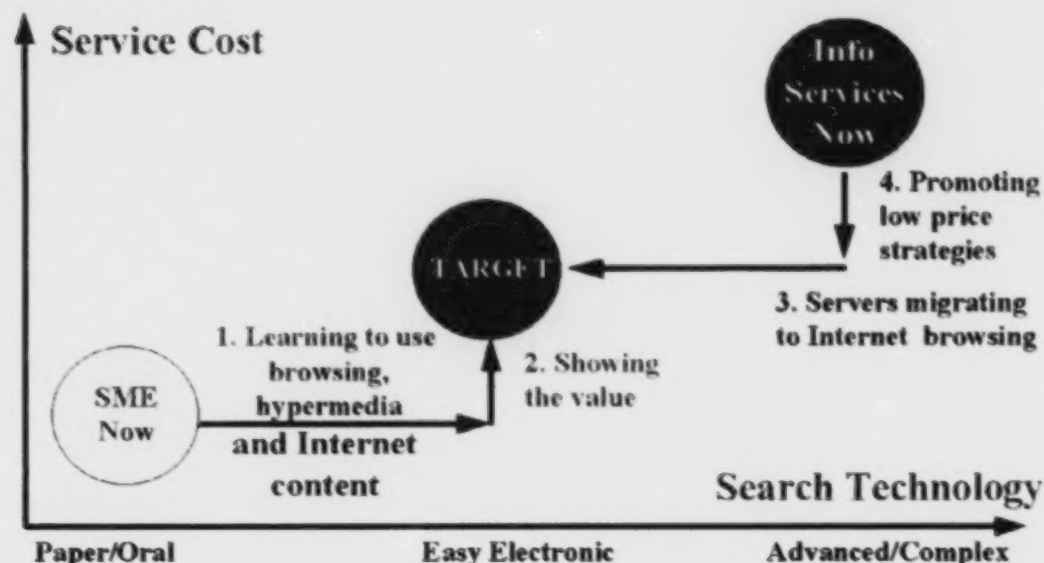


Figure 1: Filling the gap between SME requirements and online services.

Of course, awareness raising and training of SMEs as well as end-users is a must. No product like can be sold without marketing. However three other steps are required from the supply side: (1) migrating to the Internet; (2) lowering the price for browsing of information and (3) lowering the communication costs.

In the next paragraphs we shall focus on the last aspect.

3. Telecommunications means and costs

3.1. Bandwidth requirements

Ten years ago, a 300 bit/s connection was not unusual. Now most modems have at least 14,400 to 28,800 bit/s. Internet browsing is prompting a new demand for ever more bandwidth as downloading of images and full text articles is now common practice. Low speed access to Internet often results in wasting of time and every minute of work costs a company about one US dollar. So it is necessary to provide employees with high speed access to the Internet to save money. For a group of users on a LAN a 64 KB/s connection becomes a minimum for two simultaneous users. That means 24 users using the Internet half an hour a day. As one can see, a medium sized company will very soon need to have a multiple 64 KB/s connection (fractional T1). For this reason it is common to have US firms using T1 connections of 1.536 Mb/s.

Bandwidth is not cheap. In particular in Europe only very large companies are using 2 Mb/s lines internationally. Many Internet service providers (ISPs) have only a few Mb lines to the US and only sub-Mb lines to other European countries. The cost and possibly the shortage of bandwidth in Europe is what can be seen as the stranglehold of telecommunications. Let us try to understand why.

3.2. Leased lines costs

In most European countries, the current offer of online services and related telecommunications methods appears to be lagging behind seriously compared to the US. The basic costs of the telecommunications infrastructure in Europe are still priced much higher than in the US. Traditionally online services used packet switching

(X25) or PSTN to connect their clients. The cost of bandwidth on X25 networks is very high. With the Internet, access providers use leased lines (digital telecommunications lines) to connect to the Internet backbone. As a result the cost of leased lines becomes a major issue.

In the US, leased lines are now increasingly digital. There are three main categories: DS-0, DS-1 and DS-3. A DS-0 offers 64 Kb/s, a DS-1 (or T1) brings 1.536 Mb/s and a DS-3 (or T3) runs at 45 Mb/s. A DS-1 is made up of 24 DS-0 circuits and DS-3 is made up of 28 DS-1s or 672 DS-0s. There are variations build up in 64-Kb/s increments.

In Europe, telecommunications operators offer generally 64 Kb/s, 2 Mb/s, 34 Mb/s and sometimes 140 or 155 Mb/s lines.

Telecommunications operators typically price leased circuits on the basis of bandwidth and distance. There is an installation charge (1-10 months of rental), a base amount per month and a distance-based charge per month. There are no additional fees for usage. In the US the mileage portion of the pricing typically breaks down into mileage zones, such as 0-100 miles, 101-500 miles and over 500 miles. However, a single line is mostly made of several segments from two to three telecommunications operators: the long distance operator and the local telephone companies at both ends. On top of the long distance charges there are access charges that may nearly double the cost of the line (see for instance Heywood & Gronert 1996).

In Europe the price structure varies from one country to the other. International lines are priced higher. Price information is also difficult to find. There is one exception: Deutsche Telekom has its rates on the Internet. For Belgium we could use the model provided by Belgacom. We found also some information in the Tarifa study which was made available to us by the European Commission. Eurodata Foundation in London is a PTT organisation that specialises in tariff studies and is cited by many authors.

Figure 2 provides a tariff comparison for the lowest 2 Mb/s lines inside a country. The cost of bandwidth in Italy is 15 times higher than in Sweden. Belgium is four times more expensive than the US. France or Germany are over two times more expensive. The US prices presented here are lower than those cited previously. Prices in Japan are also at a high level.

2 Mb/s lines over 100 km

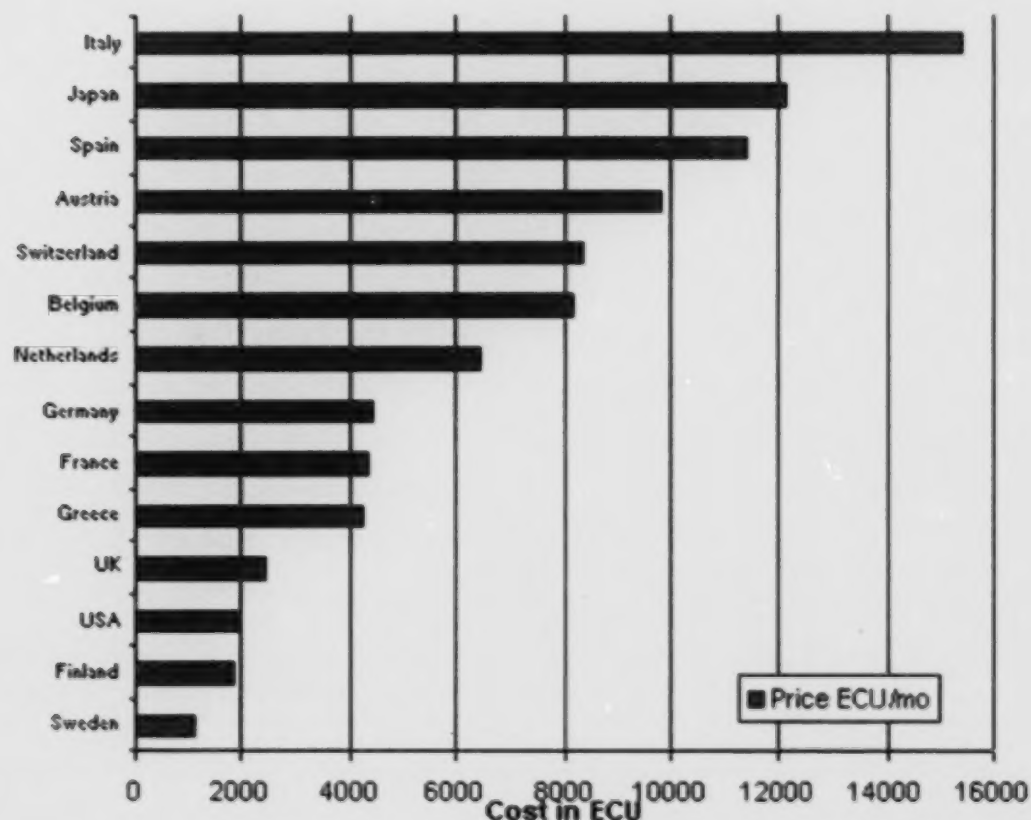


Figure 2: Monthly costs of 100 km, 2 Mb/s national leased lines in several countries.

International leased lines in Europe are even more expensive. According to BTG, the telecom user group in the Netherlands, in 1995 a 2 Mb/s leased line that crosses a border between the Netherlands and Belgium was 3.24 times the cost of a leased line with same capacity and same distance within the Netherlands (see international leased lines pricing overview of Eurodata foundation). The Tarifa study shows that a 250 km 2 Mb/s line in France costs 6897 ECU (inclusive taxes) whereas a connection to Belgium which is 300 km away from Paris would cost 22,115 ECU, or 2.67 times more per kilometre.

To understand the determining factor better one has to study the distance based charges. This is done in Figure 3. The prices of many connections in the US and in Europe have been assembled into a single picture where the cost is related to the distance. Only national leased lines are considered in this case. Figure 3 presents prices for different connections coming from different sources that have been extrapolated to identify the underlying distance charge.

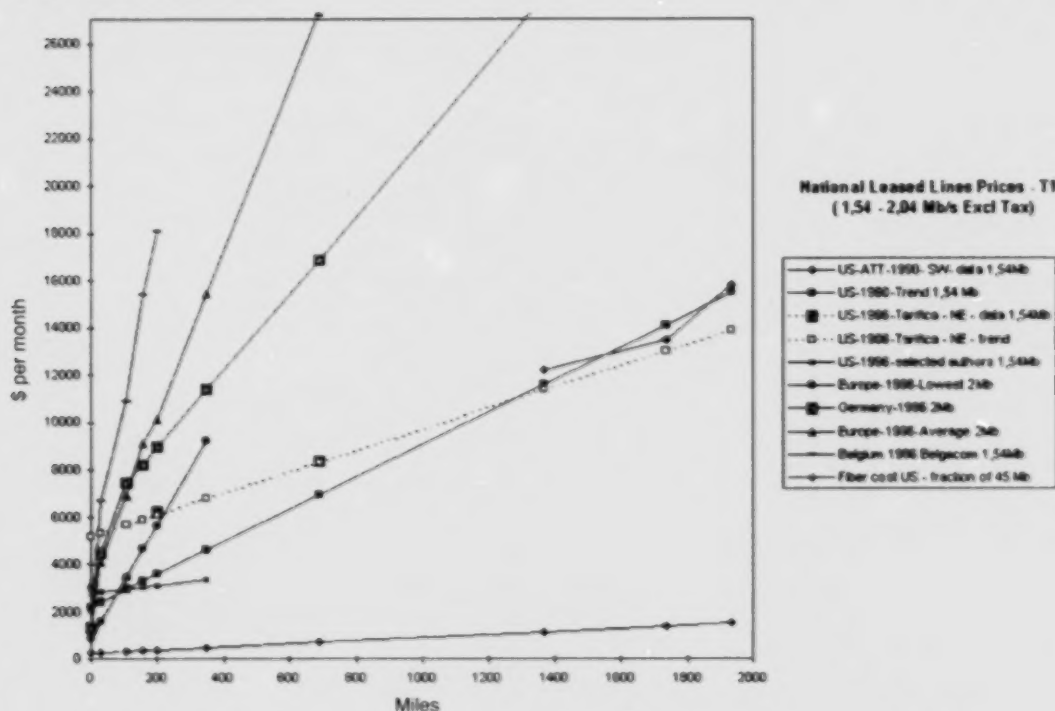


Figure 3: Comparison of leased lines costs in Europe and in the USA in relation to distance.

As appears in the figure, several authors and studies present very different price levels. The prices given for California seem to be lower than the one for the Northeast. 'For example, a DS-3 circuit between Los Angeles and San Francisco (approximately 350 miles) would typically cost about \$350 per month for a DS-0, \$3300 per month for a DS-1 and \$32,000 per month for a DS-3' (Henderson 1996). In a study by Tarifa for the European Commission, the price of a 200 miles 2 Mb/s connection between Washington and New York is priced at ECU 4901 (\$3,860), a lot more than the price given before for 350 miles. The differences can be explained partly by taxes (the Tarifa study includes taxes), different bandwidth (T1 is 1.536 Mb/s whereas in Europe the standard is 2 Mb/s), long distance operator tariffs, and access charges depending on local operators at both ends of the line.

At the bottom of the figure we have included the prices of a T3 line given by Brian Kahin in 1994. That cost is reduced to its T1 equivalent by dividing it by 28. This is approximately the cost of the fibre in the ground. The bandwidth price on a T1 appears to be 3 to 60 times higher than on a T3.

Another conclusion is that the distance charge is much lower in the US. The German prices, for instance, are 2.5 times more expensive for 1000 miles. If it would be an international line in Europe the price would be 7.5 times higher with the German prices. These are nevertheless below the average prices in Europe.

The distance charge for a T1 was \$6.8 per mile and per month in the US in 1989 for the South and the West. In the Tarifa study, which already presents high US prices, the cost per mile comes down to \$ 4.5. Kahin (1995) refers to a distance charge of \$1.6.

In Europe, according to our calculation, the average price per mile and per month is \$35. Belgium manages to have \$69.6 whereas Germany has \$16. Mercury has one of the lowest tariffs in Europe due to a low base cost but the distance charge results in \$24 per mile.

The distance charges and the interconnection cost for leased lines in Europe put the bandwidth for pan-European services at a very high level.

3.3. Internet access costs

Using the costs of leased lines we could calculate the cost of Internet bandwidth. With different connections in Europe and in the US we can model the cost of the Internet, or at least part of it. This allows us to calculate what should be the cost of the bandwidth for the users. We distinguish two kind of users: a corporate LAN with many users pumping from the Internet and a small server providing information to the Web.

Connecting a LAN with a 64 Kb/s access costs 2.5 time more then in the US. In the longer term, with decreasing costs of 2 Mb/s lines the European cost may come down below the current US price. If one looks at a 2 Mb/s access even with decreasing costs, the European costs remain much higher then current US prices. This is due to the fact that we do not expect a significant drop in the international 34 Mb/s lines in Europe and because ISP are slow to move to such capacities.

In Figure 4 we compare the connection costs for a small server which could be set up by a SME for its clients. This server connects to the ISP with a 64 Kb/s line. We analyse the cost of the bandwidth in relation to the capacity used by that ISP: 64 Kb/s lines, 1.5 Mb/s or 45 Mb/s (not including the leased line to the ISP).

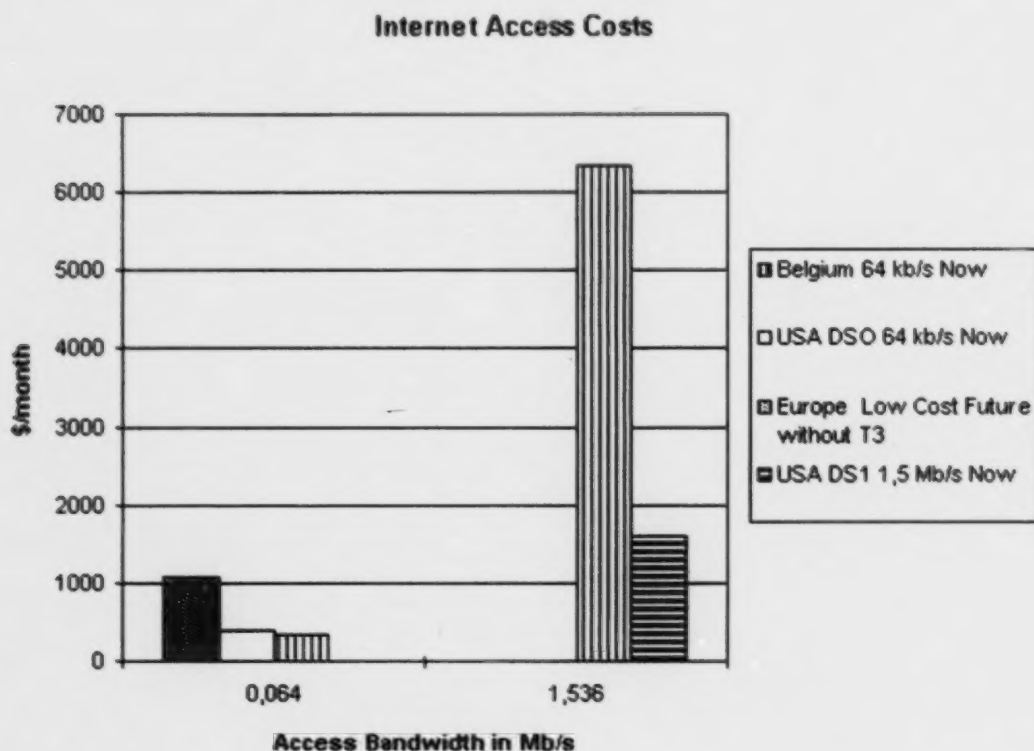


Figure 4: Comparison of Internet access costs in Europe and the US.

Small Server Bandwidth Connection Costs (64 Kb/s access)

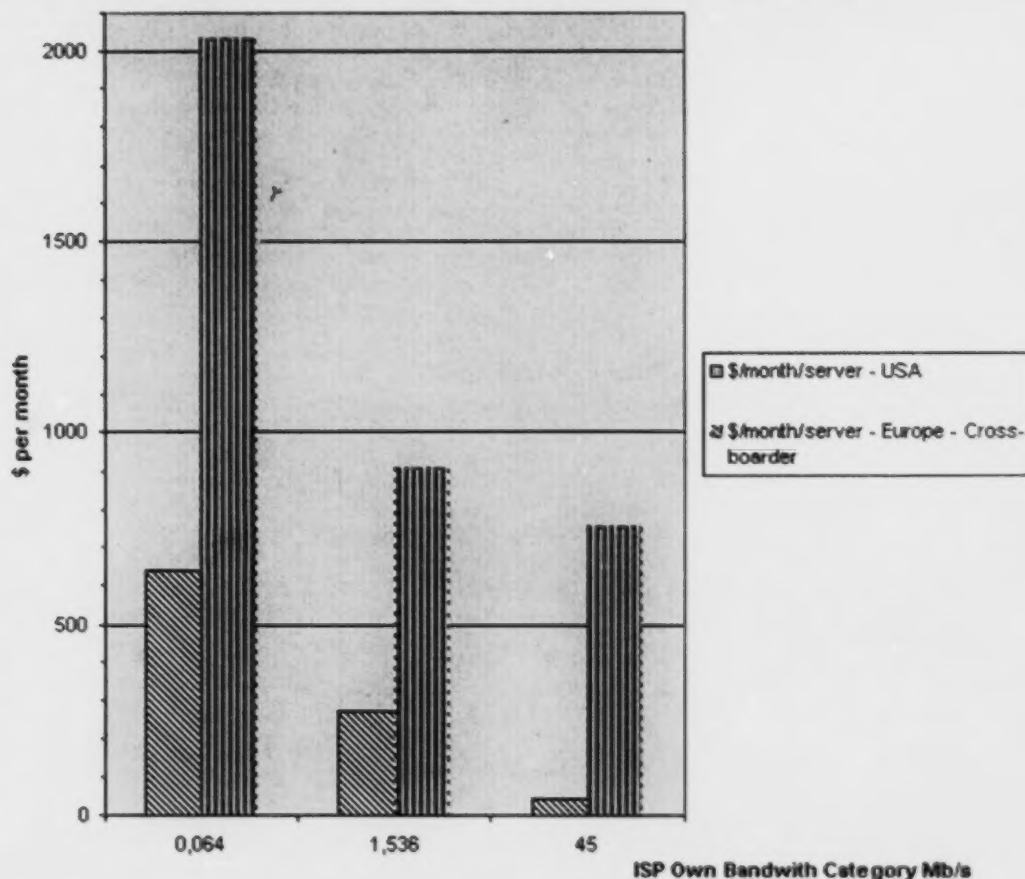


Figure 5: Impact of leased line costs charged to ISP on server access costs.

The US figures show how a cheap T3 infrastructure allows cheap access by servers to the Internet. European prices come down to \$1000 a month as ISPs move to 2 Mb/s lines. They are kept high due to the international access charges and to the distance charges. Also a European ISP must have access to the US part of the Internet with much more transatlantic capacity than his US counterpart needs. For a 64 Kb/s connection, the European costs are still 25 times higher than the US for the same capacity. US small companies can implement a small server on a T1 for a much lower price than European firms.

3.4. The future

In the future telecommunications will depend increasingly on the availability of TCP/IP networks with enough bandwidth at low prices. This may be achieved through competition after 1997. New operators such as Hermes, the company set up by the railways, offer an opportunity. Prices are already expected to decrease by 30% in 1997. However more is needed. The cost of 2 Mb/s lines on long distances over Europe should drop by almost 90%. The cost of fibre networks in the ground shows that this should be possible.

4. Impact on business

4.1. Business users

Less bandwidth and higher costs results in slower Internet access. For instance, downloading a document of say 100 KB (text and a few images) may take half a minute or even over a minute with a slow connection. In the US for the same cost per user in a company connected with a T1 it could take 1.25 seconds. It may thus take at least 2 to 3 times longer in Europe to use the Internet at the same cost. No wonder that many users complain that it is too slow.

Costs per user are high in Europe. For instance a company with 240 Internet users in Belgium would typically have a cost of \$47 per month per user. It can now be as low as \$5.5 in the US with a T1 for 288 employees connected. This shows why leased line costs should drop by 90% in Europe.

Similarly it is much more expensive to bring an Internet or an intranet server online in Europe, so they will be connected with slower lines and provide users with lower access quality. Most non-dedicated SMEs will find it too expensive to jump into the multimedia information age. All European service providers are put in a weak competitive position.

To cope with this situation, in the near future sharing of infrastructure (2 Mb/s access) seems to be the only way to have cheap access to fast Internet for the end-user, the SMEs and the small online services. Therefore the concept of small and large teleports may soon become more attractive. Local authorities may find new ways to attract business.

4.2. Multimedia technology

The status of the telecommunications system in Europe, in particular the cost of bandwidth on long distances, delays the rapid take-off of multimedia technology by European business. It is therefore not surprising that innovation and new products appear first in the US. Europe has spent large amounts of money on information technology but failed so far to provide business and citizens with cheap bandwidth to use that new technology. It is time to make sure the new policy fosters not only the technology but mainly its use. Telecommunications regulations and competition can play a role. In particular it is necessary to bring prices in line with costs and to limit access charges to reasonable amounts. Strong political will is needed to enforce these regulations properly.

4.3. Online publishing and service providers

The European environment that we have analysed involves many issues regarding information providers and Internet/intranet service providers. Information providers must embrace the Internet fast. No other infrastructure is likely to offer global bandwidth at a low price in the coming years. However in Europe they should join forces to stimulate the creation of a pan-European TCP/IP infrastructure with enough bandwidth and low, globally competitive prices.

Information providers will also have to adapt content and pricing, strategies but this is not in the scope of this paper.

Internet service providers are clearly at the core of the problem. However, as they do not own the lines, they have little power to bring T3 prices down. It is even likely that telecommunications operators (existing and new ones) would become the major Internet bandwidth providers as they own the bandwidth and know how to send millions of invoices. Internet service providers may have a strategic strength if the bandwidth is supplied by large operators with little local presence and invoicing power, such as Hermes.

5. Conclusion

Browsing information on a network like the Web is the best way for end-users and SMEs to access information. However in Europe the current high cost of leased lines is a stranglehold on the way to the information age. Costs of Internet access per user can be up to 10 times higher in Europe than in the US for the same capacity. As we have seen this results mainly from the high cost of bandwidth on long distances, in particular when cross-border connections are needed such as for Internet.

There is some hope on the horizon however: competition might bring what is needed. New pan-European networks such as Hermes from the railways and the interconnection of cable TV networks are interesting opportunities. Telecommunications regulations must bring prices in line with costs and limit access charges to reasonable amounts. Strong political will is certainly needed to enforce such regulations widely in Europe.

Jean-Claude Delcroix
President
DECADE SA/NV
Franz Merjay 117
B 1050 — Brussels
Tel: +32 2 3490190
Fax: +32 2 3490199

Bibliography

- Belgacom, Manuel d'utilisation du programme Lignes louées 1994, and programme update 1/1/96,
- Deutsche Telekom, ISDN In Germany, February 1996
- Eurodata Foundation (1996) Comparison between international leased line tariffs, Public Network Europe, February.
- Henderson, L. (1996) The price of WAN connectivity — choosing the right WAN service for your business typically comes down to pricing, Byte, <http://www.byte.com/art/9605/sec8/art3.htm>.

- Heywood, P. and E. Gronert (1996) High speed data services, *Data Communications*, March.
- Heywood, P. and To Chee Eng (1995) Global supernets: big pipes, big promises...and one big problem, *Data Communications*, 21 September.
- Kahin, B. (1995) The Internet and the national information infrastructure, Harvard University 3/1995, <http://ksgwww.harvard.edu/iip/KAHINPA.ASC.txt>.
- Stuck, B. (1990) *Telecommunications Equipment And Service Trends: 1990-1994*, William K. Woodruff & Company Inc., Dallas, TX.
- Tariff Data 01.01.96, prepared by Tarifa, a division of PBI Ltd, Stevenage, for the European Commission DG XIII.
- van den Broek, F.G.H. and Maarten Looijen (1996) Cost-effective management of international networks, http://www.ctr.columbia.edu/vi/papers/art_2.htm.



The market prospects for consumer online services in Europe

Martin White

TFPL Ltd, UK

Abstract: *The development of online services targeted at the consumer mass market in the USA is outlined and the main European services are profiled. The factors that are affecting the rate of growth of these services in Europe are described, and from this analysis it is concluded that it is very unlikely that the demand for these services is going to grow as rapidly in Europe as has been the case in the USA.*

1. Introduction

Since the earliest days of the online industry there has been constant debate over the extent to which services tailored to the market environment in the USA would be adopted in Europe. The rapid growth over the last few years of online services designed to appeal to a mass market audience in the USA encouraged many of the companies operating these services to consider launching them in Europe. At the same time a number of European companies, observing the profits generated by these mass market services in the USA, decided that the business model could be adapted to the European market, either on a national basis or on a European-wide basis.

It is still too early to give a definitive view on the success of any of these services, but an examination of them is instructive in understanding some of the fundamental differences and similarities of the market for online services in Europe and the USA. This paper concentrates primarily on the European market and does not attempt to give a comprehensive picture of market and industry developments in the USA.

2. Categorising the services

The characteristics of the services to be covered in this paper are:

- they deliver a broad range of information, entertainment and transactional/e-mail services;
- they are targeted at individual users with no previous experience of using online services;
- they are often billed to a credit or charge card, on a monthly subscription basis.

Although the services are primarily targeted at the consumer market, the small office-home office (SoHo) market is also an important sector. Additionally in the USA there are a substantial number of successful services used by individual investors and these have not made any attempt to enter the European market.

Within this broad definition the services themselves fall into three categories:

- US-owned services which are available and supported in Europe, for example America Online (AOL);
- European-owned services marketed in more than one country, for example Europe Online;
- European-owned services marketed in one country, for example UK Online.

The use of the word 'marketed' is important because there are few barriers to accessing these services world-wide over the Internet.

3. US-owned services

The origins of the USA services go back a very considerable way, with CompuServe and The Source both beginning operations in the mid-1980s. GEnie was launched by General Electric Information Services in 1985, and Prodigy was launched by IBM and Sears in 1988. The Source was acquired by CompuServe in 1984. America Online also dates back to 1985, as Quantum Computer Services. All these services were targeted at the bulletin board user in the USA, which was a large market that had grown up as a result of low-cost communications, a large installed base of PCs, a keyboard literate user base and the ability to communicate in English across a continent.

In the late 1980s the four majors — CompuServe, America Online, Prodigy and Genie — all grew at roughly the same rate. By the turn of the decade it was starting to look like a two horse race between CompuServe and

Prodigy, with CompuServe having very good e-mail services and Prodigy concentrating on the games market. Then Steve Case, who had worked his way up the marketing ladder in AOL, was appointed Chief Executive Officer of the company in 1992. Case then started to market the service in some highly effective ways, and by 1994 AOL had overtaken Prodigy to take second position to CompuServe.

During 1994 and 1995 two more services arrived on the scene. In late 1994 Apple launched eWorld and in mid-1995 Microsoft launched the Microsoft Network (MSN) as part of the Windows 95 package. eWorld was developed in conjunction with America Online and offered users a very innovative front-end. The main claim to fame of MSN was that it could be accessed directly through an icon on the Windows 95 desktop, a facility that resulted in a complaint to the Department of Justice by many of Microsoft's competitors, although in the end no action was taken.

Table 1 summarises the rate of growth of subscribers to the leading services over the last five years, in thousands of subscribers.

Table 1: Rate of growth of subscribers to the leading services over the last five years, in thousands of subscribers.

	1990	1992	1994	1996
America Online	109	200	1500	5500
CompuServe	740	1130	1500	4700
Prodigy	440	1000	1200	1450
GEnie	70	100	100	55
eWorld			65	130
MSN				1000

The main issue that needs to be addressed here is the extent to which subscriber numbers are any indication of commercial success. There is no doubt a very substantial number of users who have logged on to, and registered for, these services but who have not made any substantial use of the services once the initial trial period had elapsed. It is also very likely that users subscribe to more than one service. Estimates of the level of what is often referred to as 'churn', i.e. subscribers leaving one service to move to another service, vary considerably but it could be as high as 30%.

America Online in particular has also come in for adverse comment from the US Securities and Exchange Commission for the way in which the company treated many of the costs of acquiring subscribers as a capital expense, and so writing off these expenses over time.

4. The Internet — opportunity or threat?

All the US services built their base of subscribers in a pre-Internet environment. Certainly the Internet was widely used in the USA from 1992 onwards, but the proprietary interfaces that the consumer services offered were much easier to use than Mosaic and the other browser applications available at that time. The introduction of the World Wide Web and the Netscape Navigator browser changed the market dynamics in 1995. The question for the various services was whether they should embrace the Internet or compete with it. After a number of dismissive statements from senior executives, the rush to adopt WWW standards gathered considerable momentum in late 1995 and 1996, with only America Online retaining its proprietary platform. Two of the major losers were AT&T, who had bought the Interchange software from Ziff Davis, and Microsoft, who totally underestimated the rate of growth of the WWW.

5. European-wide services

During 1994 the attention of the US service providers switched to the European market. CompuServe had been quietly building up a strong subscriber base in Europe, mainly through the provision of a good e-mail service. Without going into the complex pre-launch histories, by the end of 1995 it was clear that there would be three competitors to CompuServe. America Online announced a deal with Bertelsmann (probably the world's largest media company) to set up AOL operations in Germany, France and the UK. Microsoft's ambitions for MSN were always global and the company put quite a lot of effort into obtaining local content.

Europe Online was established by a diverse range of shareholders, brought together largely by Candice Johnson who had been involved in the launch of the Astra satellite television service. The theory was that there would be publisher-shareholders in Germany, France and the UK, together with support from the Luxembourg financial community and AT&T, which was contributing its recently acquired Interchange software.

The first to launch was MSN, followed quickly by AOL/Bertelsmann in Germany and then AOL in the UK. AOL

France was launched at MILIA in February 1996. In both Germany and the UK the AOL marketing machine has been very impressive, with (for example) deals being done to give schools low cost access to AOL in the UK. The key marketing route for AOL in Europe, as in the USA and also as mirrored by CompuServe, is to attach disks containing the registration software to magazines on display in newsagents, often referred to as cover-mounting.

By comparison Europe Online have been very low-key, a result of having to move from the Interchange platform to an Internet access route just before launch in December 1995 and apparently considerable differences of opinion between the shareholders about the direction of the business. The end result is that Europe Online has crept out, rather than launched.

Karel van Miert, the Commissioner responsible for competition policy in the European Union, initiated a review in late 1995 of the extent to which the AOL/Bertelsmann/Deutsche Telekom and Europe Online services are contrary to competition policy. The report was originally scheduled to be published in April 1996 but as of June 1996 it had not been released.

6. National services

A number of national consumer online services have been established in Europe, mainly in the UK. UK Online was set up in 1995 with support from Olivetti, which also owns Italia Online and provides services targeted at families, including games and educational content. Tel-Me, a service operated by Phone Link, provides a service of more interest to the business community with detailed maps, Yellow Pages and travel information, for example. To overcome concerns by users about telephone line charges the service provides subscribers with a substantial proportion of the database on the PC hard drive, with the indexes and updated information being accessed over the network. When the subscriber is reviewing the material the service logs off until the next search is requested.

In France there is already a mass market service, Teletel, but this was built around a dedicated terminal and a rather basic character set. Faced with potential competition from Internet services, and from AOL and CompuServe, France Telecom moved quickly in early 1996 to lower the cost of the Teletel services and to provide Internet/Teletel gateways. In addition the Wanadoo service, effectively a Yellow Pages service, was launched in early 1996 and Infonie is an entertainment service provided by Infogramme, a multimedia publisher. In the autumn of 1996 Fun Online is scheduled to become available in Germany, providing interactive entertainment for children based around a subscription to a CD-ROM which provides the backdrops to the games.

7. Factors affecting the adoption of online services

There are many factors that are likely to affect the rate of growth of subscribers to these services in Europe and the extent to which they will be used.

7.1. Installed base of PCs and modems

There are 149 million households in Europe, compared to 97 million in the USA, and so at first sight the market for consumer services would seem to be very substantial. However the penetration of consumer electronics equipment into European households is substantially lower in Europe.

Historically modem prices have been very much higher in Europe than in the USA. This situation has now changed but the costs of obtaining type approval in each country, and of supporting modem distribution, have still meant that modem prices have remained high.

7.2. Teletext

Broadcast teletext services have been widely available in Europe for over two decades. These services are usually operated by the public broadcasting organisations and use the vertical blanking interval in a standard television signal to transmit a sequence of index pages and information pages. The graphics are rather limited but the services are provided as a free public service. Virtually all current television sets have a decoder incorporated. The range of information on these services is quite wide, including news, weather, sports results, TV listings, travel information and other leisure information. The technology is also used to provide subtitling for deaf viewers. It is also possible to provide a limited degree of interaction through the use of touch-tone telephones, and in the UK there have been some trials of online shopping through teletext services. Because of the localisation of the commercial television companies it is feasible to produce local advertising and news services. One recent development has been to put personal messages and advertisements on the service. One of the most successful areas for UK advertising on the teletext services has been that for holidays, and Teletext claims that over 20% of all holidays are sold over this network. This is an area that the BSkyB Intertext service is also planning to develop. These teletext services represent a substantial challenge to the consumer online services that do not exist in the USA. Instead of having to dial in to an online service for sports results, weather, railway timetables and news, teletext services offer all this free without moving from the television.

7.3. Telecommunications costs

Compared to the situation in the USA the cost of accessing online services in Europe has always been high

because of the tariff policies of the telecommunications operators. In the USA local calls are free and time-independent, and that is often all it has taken to access one of the major networks to gain access to the services. In Europe the options have been to have dial-up access over the public switched telephone network (PSTN) or to use dial-up access to the public packet-switched data networks (PSDN), both of which are substantially more expensive than in the USA. Another issue that has to be taken into account is the ease with additional lines can be installed, especially in the domestic market, to overcome the problems of extensive use of online services effectively cutting off the household from voice telephony.

7.4. Value added tax

One of the differences between the US market and the European market is the level of direct taxation on online services. Tax levels in the US are usually around 5 to 7%, but in Europe they can often be twice that level. For the business user this tax can be passed on in terms of goods and services to customers, but consumers do not have this option and (to take the case of the UK) pay 17.5% value added tax on both online subscriptions and on telecommunications charges. Comparisons between services need to take into account whether or not VAT has been included in the pricing.

7.5. Language

One of the most difficult problems to overcome in the development of a European online services business is that of language. Even within national boundaries there may be more than one language in common use, with Belgium and Switzerland being the best examples of this situation. Europe is fragmented into 40 distinct languages and language variants. Within the European Union native speakers of German (78 million) outnumber those of English (56 million) since the unification of Germany. As far as the online industry is concerned the problem is not only which is the first language, but to what extent will users who do not have English as a first language be prepared to use English-language online services. Statistics on second-language capability are difficult to interpret as the degree of skill in the second language can vary widely. It is possible for someone to be able to read an article in French offline with the aid of a dictionary, but that same person may be very uneasy about using French language commands and search terms. For business services users may be prepared to use other languages, but this is very unlikely to be the case in the consumer market. However the use of English is clearly going to increase over the next few years, and the use of Internet access to US services and Web sites will do much to stimulate this growth.

7.6. Information culture

Another important variable to take into account is the different 'information cultures' within each country. This is a rather broad term to cover the use made of daily, weekly and regional newspapers, readership of professional and consumer magazines, the use of book clubs and the development of public libraries. The demand for news is being satisfied in very different ways in Europe, and this has an implication for the extent to which online news services will be able to compete with, or complement, the newspaper industry.

7.7. Electronic mail

By 1996 it is expected that more than 80% of large companies will have access to the Internet and there is expected growth among small and medium sized companies. Mail is the simplest and most widely used network service, both on the Internet and on other networks. Online providers report that communications services, which include e-mail and the Internet, are among the most popular application. One of the most profitable elements of the service range of the online services in the USA is that of electronic mail.

It is still not clear just how strong the market demand is for e-mail in Europe. In relative terms the postal services work quite effectively, and it is difficult to see how heavy the use of electronic mail will be in the domestic market. Families in Europe tend to be more closely situated in geographic terms than in the USA, and in any case the market would be more limited than in the USA by reasons of language and to a certain extent a fear by older adults of computing. This reticence to use computers will of course gradually subside as the average age of users decreases and their computer literacy increases.

7.8. Electronic shopping and banking

Any discussion of electronic shopping had to take into account issues concerning advertising and product design in the various countries of the European Union. Although many of the barriers towards the establishment of a Single European Market have been removed there are still some significant restrictions on cross-border advertising. Apparently similar products may in fact have variations to meet regulatory requirements or consumer requirements. For example, electrical equipment with fitted plugs for the German market cannot be used in the UK market without changing the plugs. This may not be immediately apparent in the product description. Another complicating factor is that trademarks are issued on a national basis, and so it is entirely possible for a company to contravene the use of a national trade market through advertising the product in that country. Customer support for faulty goods, or the return of goods not meeting the product specification, also would need to be set up on a pan-European basis. The result will be that for some time there will be only a limited demand on a European basis for goods ordered on electronic shopping services, unlike the situation in the USA.

On the other hand, electronic banking has become very popular in many countries of Europe, in particular

Germany and France using the public videotex services, and in the UK using voice telephony but with increasing use of PCs. It is likely that accessing financial services of all types is going to be more heavily used than electronic shopping, especially now that it is possible to buy insurance policies on a cross-border basis. Many of the large insurance companies have subsidiaries in other European countries to a much greater extent than banks.

8. Critical success factors

Taking into account the market development issues outlined above, it is possible to identify some of the critical factors that need to be taken into consideration when assessing the market potential and business prospects for mass market online services in Europe.

8.1. Market positioning and service differentiation

The services currently available seem to be hedging their bets as to the most attractive market to attack. The problem is that the business community is currently the most attractive target market as it is likely to buy multiple services, and to worry less about VAT and telephone call costs. However the consumer market, because of its size, is also a very attractive market but growth initially may be rather slow. The result of this is that most of the service providers are offering a range of database content, typified by CompuServe with everything from games and reference information through to premium services aimed at the business community.

The situation in Germany is especially interesting. Information from Deutsche Telekom seems to indicate that the recent growth in the subscriber base of T-Online has been from the consumer sector, and yet in the reallocation following the recent joint venture between Deutsche Telekom and AOL/Bertelsmann, T-Online is going to be the delivery channel for business information.

The AOL approach is that it wants to attract anyone with a PC, a modem and a credit card, which gives a very large potential market but also means that the company has to have content to keep this market satisfied. EOL seem to be adopting a much more targeted approach, looking more for high-income earners and business users whose working environment includes the home.

This leads into the contentious issue of subscriber counting. The mobile telephone industry has already learned, to its cost, that it is not the total number of subscribers or the number of new subscribers that are important metrics. The key metrics are revenues per subscriber and net gain in subscribers after allowing for cancellations. AOL in particular is very aggressive with publicising subscriber numbers, but there is a difference between the many subscribers who have been given a free disk of access software and some free time, and how many of those subscribers are using the service intensively in six months time. The figures quoted by Bertelsmann suggest a revenue per subscriber target of 200DM per month. This is equivalent to a usage level of around two hours a day each month, which is in line with current levels of US usage, but whether this is realistic in Europe remains to be seen.

8.2. Pricing and billing

Just how price sensitive this market is going to be is still far from clear. The strategy in the USA seems to have been to keep the monthly subscription steady at around \$10 per month, and then to change the number of free or discounted hours and the premiums for special services. If Internet access is bundled in then this also has been taken into account, which is one of the problems that has affected Microsoft in Europe. As has already been mentioned, in the USA the effect of sales tax is quite small and can almost be discounted in any comparisons. That is not the case in Europe where the value added tax issue could be quite a substantial one in marketing terms.

Billing is also going to be a difficult problem, as the number of currencies is almost certainly going to be greater than the number of languages supported. Even for businesses this is going to be an issue as the costs of raising small-denomination cheques or bank drafts can be larger than the value of the check. For a consumer a credit card or a charge card may be the only option. In the UK credit cards are very widely used but this is not the case in Germany, and in Scandinavia in particular.

There is also an issue in the business market regarding corporate charge cards and credit cards. The corporate card, where all the expenses are reimbursed with only the minimum of due diligence, is not widely used in Europe and many companies will only reimburse expenses incurred on a personal card, or for that matter a corporate card, when there is a receipt. Direct debiting of services from a credit card will only show up in the monthly statement. Should users go over their credit limit by a major purchase in the month (for example a holiday) then the service operator could have some interesting problems collecting the monthly payments!

Another aspect of this issue is the extent to which the card issuers would be prepared to be liable for any problems experienced with the service. In summary, reliance on credit cards may give rise to some complex business problems which are virtually unknown in the US market.

8.3. Marketing

Raising the profile of their online services is going to be of extreme importance to service providers, and it is the balance of the media spend that is going to take time and vigilance to optimise. Some of the techniques that will need to be evaluated are:

- co-marketing with PC manufacturers and distributors;
- cover-mounting disks onto magazines;
- advertising;
- co-marketing with publishers and media companies;
- exhibitions;
- editorial coverage.

The marketing budgets of the major services are very substantial, with CompuServe planning to spend over \$180 million in 1996 on new subscriber acquisition.

8.4. Content availability

One of the most important differences between the mass market services and professional online services is the percentage of gross revenues paid to content providers. The mass market services, exemplified by Microsoft, are giving the content provider the majority of the gross revenue, often more than 60%. For certain key databases (such as the Helicon Encyclopaedia on CompuServe) the service provider may be willing to pay to ensure exclusivity.

Working out mutually satisfactory agreements is going to be a growth market for the legal professionals in Europe. There are no precedents to work with and the experience of the US market is not applicable. In the early stages of marketing a new service the investment by the provider in terms of hardware and personnel infrastructure is going to be very considerable, and so setting out the terms of a deal that may need to last for several years is clearly going to be very difficult. On the other hand, there may turn out to be only a very limited market for a particular database and this may not be obvious for some time.

In the early stages of market development few content providers are going to want to sign exclusive agreements unless very large sums of money change hands. One problem that this gives rise to is that many sources appear on more than one service, such as the major press agencies and company information providers in the business sector, and radio/television programme listings and travel companies in the consumer sector. This is going to make service differentiation very difficult, especially where a content provider also has a good Web site.

One of the unresolved issues in Europe is the extent to which users in both the business and consumer sectors will use these services to access news. Readers of newspapers look for more than just 'news' and tend to browse through a paper. There is certainly a place for news services in an electronic format, but is it unlikely that they will represent a 'killer application'.

8.5. Defining the basic services

In the USA the core services include electronic mail, news, weather, sports news and investment information. The forums and chat groups are also very popular. Electronic mail in particular has been very profitable for the leading service providers and the US newspaper industry has embraced the online services with a substantial degree of enthusiasm, but the industry probably has less to lose from loss of advertising revenue than in Europe. It will be interesting to see what comes to be regarded as core/basic services in Europe, as this will be the pricing/service differentiation battlefield for the next few years. Although at the early stages of being a basic service, it is likely that the important differentiation will be the ease of setting up and administering World Wide Web home pages using a simple HTML compiler. CompuServe has already been quite successful in this regard.

9. 1996 — the story so far

The early part of the year saw an immense flurry of activity, with deals being done almost daily in an effort to build Internet capability and WWW browsing capabilities into the proprietary interfaces. Apple, beset by financial problems, closed down eWorld and UK Online lost two senior staff. Prodigy was sold off to a venture fund, to the considerable relief of both IBM and Sears, and MSN was effectively relaunched as a WWW service despite the considerable investment made by information providers developing content for the original proprietary platform. GEnie has also been sold to a venture fund. In June CompuServe announced a rapid migration to the WWW as H&R Block, its parent company, started the process to sell off CompuServe, as well as launching Wow! as a service targeted solely at the consumer marketplace in the USA. America Online also had its problems in June 1996 when it announced a pricing reduction, and this announcement combined with fears about the future of AOL in the Internet world resulted in a sharp drop in the share price.

In Europe all the services are struggling to maintain their ambitious growth targets. In June AOL reported that the service had 150 000 subscribers in Germany, with the announced target being 800 000 for the end of 1996. CompuServe had nearly 800 000 subscribers. Europe Online reported that the service had 25 000 subscribers in Europe. The announced target for the end of the decade was 3.3 million users. Subscriber numbers for the UK services are difficult to establish but Tel-Me has 6000 subscribers.

At this point in time the future for these services does not look very promising. CompuServe has the benefits of an established, and quite loyal, subscriber base, and AOL has considerable financial resources. The questions over the shareholding, and thus direction, of Europe Online remain unresolved and all of the national services are struggling.

Online Information 96 Proceedings

Page 318

It looks as though the rapid growth in the use of the WWW in Europe before the services could become established, the existence of free public teletext services, the costs of telecommunications and a rather different information culture are just some of the reasons why all the US services, and the European-owned services, have not seen rates of growth to match those in the USA.

It seems unlikely that there is a sufficient market in any one European country for more than two mass market services, providing they provide national language content and support. The question that is starting to be resolved is will these services be pan-European (i.e. AOL/CompuServe/Europe Online) or is there still a chance for a strong national service, such as UK Online or Wanadoo. By the end of 1996 it will be somewhat easier to predict the shape of the online industry but one thing is already fairly certain, and that is that the lessons learned in the USA are of little help in developing the European market.

Martin White
TFPL
17-18 Britton Street
London EC1M 5NQ
UK
Tel: +44 (0)171 251 5522
Fax: +44 (0)171 251 8318

Acknowledgement

I am grateful to my colleague Mary Coyne for her assistance in the preparation of this paper.

Reference

White, M. and M. Coyne (1996) *Mass Market Online Services in Europe*, TFPL Ltd, London and New York.



The generation gap: Minitel in the face of the Internet

Philippe Gilbert

Computer Answer Line, France

1. Introduction

1.1. Company presentation

Computer Answer Line (CAL) is a software company based in Lyon, specialised in enhanced online services and a member of the World Wide Web Consortium. CAL is the provider of SIAM technology and the technological leader in providing new generation French services.

1.2. History

Starting in the 1980s the online world has seen two models:

- the French model (Minitel), based on a 'dumb' terminal but with some strong advantages: wide free distribution, access to universal services (telephone directory);
- the American model (Internet) based on services dedicated to computer freaks. For a long time CompuServe had a reputation of an online service for computer scientists but with the strength of a large market enabling a financial success. In France, we may distinguish two periods after the initial launch and first period of growth:
 - From 1985 to 1990 the services provided became more and more professional, the number of users grew and the network infrastructure offered a wide range of services;
 - From 1990 to 1995 we saw a decline of the Minitel, mainly for three reasons: service providers (SPs), after the growth period, estimated that they were able to increase prices; Minitel's connotation of being expensive; and a technological breakdown as users had more and more PCs and the network speed did not evolve. All strategic errors for France Telecom (FT).

2. Situation in France and overseas

At the end of 1995 the Minitel network was the biggest and oldest online service with 7 million users and 20,000 service providers. One year later, the Internet rushed into this quiet but profitable world.

Why?

If we exclude the arrival of new online services (AOL, Europe Online, Infonie) which tried to attract users, we may consider that the 1996 big bang had two main reasons:

- people saw in the Internet the possibility of going beyond FT, so a large craze for this new world attracted users, and this feeling was emphasised by the fact that FT was a state owned company and did not jump into the private sector;
- FT itself, seeing these critics, wanted to make a strategic break and decided to invade the Internet world for its online evolution. The Minitel was forgotten and we lived in an amazing period where there were at least 20 projects for launching the best online service with the greatest payment solutions and the largest information provider. If we trusted them, the information providers would spend all their resources by uploading information on all the online platforms.

The situation is now more stable and we see three kinds of actors:

- the software and service companies: Internet service provider (ISPs), software companies (Web vendors) ...;
- the online world, mainly American (CompuServe, AOL) and some French (Infonie);
- FT, which acts as an ISP as well as an online company (Wanadoo) and the network provider (its original job).

Facing these actors are the information providers (Ips). During this period, nobody moved and the main question was 'where should I go?'.

This seems perhaps a stupid question to you, but you have to know that some ISPs have a large computer investment in their Minitel service with large technical teams of engineers and online editors. The economical model of these companies was based on the 'kiosk'. For those who never studied the French market, this means you connect to an online service without any identification and registration, and one month later on your telephone bill you receive just one line between many that is your online invoice. We call this 'painless billing'.

If we take a poll today, the only thing people are sure of is the technological choice for information. We mean by this that they are quite certain that they need to use the HTML standard for their information content in order to prepare the next five year evolution. Another question rapidly pointed out: of this 7 billion French Francs (nearly £1 billion), what about the revenues? Why these questions? Because they are used to earning money with their service and they saw the spectrum of a free information world where they need to justify the information price and set up sophisticated mechanisms for user invoices. So, after this big media bang, the next step was nobody's move.

And the users? Nobody seems to pay attention to the users. Are they equipped with up-to-date multimedia equipment? Will they agree to subscribe to two or three online services and ISPs? Are they well equipped or afraid of the new economic models where they have to show their credit cards at each page?

Take a quick look at the figures in June 1996: FT's online service Wanadoo announced 17,000 users; Infonie (the French based online service from Infogrames) did not say anything (unofficial figures of 2500 subscribers); no information from AOL; and CompuServe announced 30,000 users. If we had the ISPs we could reach between 100,000 and 150,000 users. Not enough.

The first answer to users' questions comes from Minitel. After several years of stagnation or regression, Minitel traffic grew during the first quarter of 1996 and it seems that Internet mediatisation had a positive impact on Minitel use. So, is there competition between Minitel services and Internet services? The main and largest ISPs say 'No'. They estimate that they have at least five years of potential business on the Minitel and they are taking time to prepare the next generation services.

I think that either users are still not well equipped or they are afraid of the new economic models where they have to show their credit cards at each page. Take a look at the other side of the Rhine River. Deutsche Telekom (DT) had a videotex network called 'BTX' based on old technology. This network, close to Minitel principles but with more constraints, has never been as successful as Minitel. They decided two years ago to move to a modern architecture. What was the result at the beginning of 1996: more than 5000 services available and 1 million subscribers.

What are the keys of this success:

- the technology development was realised by a work group involving ISPs, technology providers and DT;
- DT helped the development of the first services;
- DT engaged itself in promoting the network, distributing a client package;
- there was in the package an open door to the Internet.

In summary, we may say that they have applied the Minitel principles to a new technology. We saw the same success in Denmark, where the telco: TeleDenmark launched a similar product called DIATEL with our SIAM technology.

So, in this liberalised business, is there a place for a public utility service and old fashioned centralised networks? Our answer is 'Yes!'

3. The new online model

Ten years of Minitel experience has modified the Internet business in France and some easy to use model and technology will modify the market. Who will suffer, who will gain and when? Last Spring, the annual World Wide Web conference was held in Paris. We saw a lot of American Internet gurus coming to us and asking: 'I want to see a Minitel, what does it look like? How do you use it?' and we're sure that ISPs who will win in this business will take some good tips from the old fashioned terminal.

Why? If I compare Internet development in the UK and France I see one main difference.

- You, in the UK go into this new business without any ulterior motive: you consider the Web as a new media and begin to see potential application development.
- We, in France, are sure of this business (in size it's as important as the TV business) but don't know how to advance with it, so, the solution is in between.

Tomorrow's online model will need to be:

- as familiar to the general public as phone or TV;
- as natural as a public utility invoice;
- as easy to use as your PC applications;
- transparent when connecting to the service;
- and of course as complete as a newspaper.

Market trends see the development of new generation networks:

- using the standard technologies but providing enhanced functions;

Online Information 96 Proceedings

Page 322

- open to everybody but giving extra services to registered users;
- giving the information provider the possibility to run both public and private services on the same platform.

The current actors, both telcos (France Telecom, Telefonica ...) and private (AOL, MSN ...), have a strategy based on attracting information providers in exclusive partnerships, which has three negative effects:

- for the information provider who is too closely linked to the network provider, in terms of the subscriber (he needs to trust in the Online service for its figures) as well as technology;
- for the users who may have to subscribe to a large number of online services to get the information looked for;
- these new online services are mainly targeted to a large public and do not give any room to small and medium enterprise services.

Based on these facts and on a need for information providers to launch independent online services, there is a need for a national and European network of online services. Who is likely to win tomorrow?

- not the world online services (AOL, CompuServe ...): their present difficulties in staying in the technological race are not a positive sign. But more probably;
- telcos (FT, Telefonica ...) as they are already in touch with a large public. They will need to separate their public utility role from their private goals;
- and also national online services if they reach a critical size.

4. And what about the users' technical requirements?

For quite a long time, all the talk about online has been concentrated on Internet and more especially on the World Wide Web. The Internet network is widely used worldwide and is growing exponentially. The Internet is the global transport network used by numerous applications and protocols (telnet, FTP, SMTP, SNMP, gopher, and so on), including the WWW HTTP/HTML transport protocols.

The WWW world consists of documents linked by hypertext links. To follow a link, the user clicks with a mouse on a highlighted word in the current document which causes the linked-to document to be retrieved and displayed. All documents are named using global addresses (called Universal Resource Locators — URL) including the access protocol used to retrieve the document and the document location in a way significant for the access protocol (for example FTP address). The system is really a worldwide one as any document could reference any other, even if stored on another system located at the other end of the world. A URL could point to a script and in this case the server will send back to the client the output of the script and not the script content: this mechanism allows the implementation simple of processing.

We don't question the Internet as the network but we think that currently there is no technology on the Net which bring an answer to the ease of use reached by old technologies both in information browsing and in service marketing. What about these nice advertisement pages with an awful address like 'http://www.alpha-beta.tm.fr/product/new.html' for getting information?

The HTML standard we find on the Web is only a presentation model and we need more, especially in terms of applications and security. Moreover, we need applications with little intelligence on the 'client' but piloted by the server to help the novice end-user.

The documents could be created very easily using HTML (Hypertext Markup Language): they are created in standard ASCII text including simple tags to specify the document title, paragraph boundaries, headings of several levels, highlighting, lists and links to other documents.

The HTTP communication protocol between the client and the server is very simple: the client establishes a connection with the server and sends a request containing the word GET followed by the name (partial URL) of the document to be retrieved. The server responds with the document contents and then immediately closes the session.

The difference between the two protocols — the new multimedia interactive one (SIAM) and HTTP/HTML — is the session maintaining the link between the client and the server.

A Web browser only retrieves a document to display (or sends a document to the server) and closes the communication. When the user clicks on another word to retrieve another document, a new communication is temporarily established for the time needed to retrieve the document.

What appears to the user to be a single continuous 'session' is in fact a succession of corresponding responses. Therefore, the WWW servers do not keep any session context and are not always well suited to implement high value added applications best supported by a bi-directional data exchange between the client and server during a single session.

It is then clear that the Web and the browsers (such as Mosaic or Netscape) are performing tools, easy to use, but they are not always well suited to the creation of high value added interactive applications that require a continuous session between the client and server such as could be implemented with SIAM — a multi-windowing, object oriented interface. By the way, Internet experts are well informed of these Web limitations and are currently trying to find other protocols to fill the gap between the capabilities of the Web and the user requirements.

Various extensions of HTTP and HTML are presently being studied in order to extend the current capabilities of the Web. For example, the Sun Java language allows the client software to run local 'operative objects' written in a machine-independent p-code. The HTTP-NG protocol has been designed to offer a session control protocol allowing a single session to be maintained between the client and the server. The latest HTTP 1.1 draft has also included a Connection Keep-alive parameter, enabling it to request a persistent connection. All these proposals keep the request/response paradigm used up to now. From the point of view of Internet experts supporting this improvement, it seems the main reason behind it is mainly to solve the performance problems caused by the setup of a new connection for every HTTP request, or to implement more easily improved security mechanisms (and not mainly to take advantage of the possibility to implement more easily complex data processing applications). The answer relies partly on the SIAM technology which has for five years now demonstrated its capabilities in answering all the these questions.

The first connection on the Web is not very easy for a beginner and once connected it is not so easy to find something without any help or without a lot of time. We think that users (the public at large, not the computer freaks) haven't the time at their disposal and it's a hindrance in the use of the Net technology. (In October 1995 there were eight million pages on the Web: six months later in April 1996 there were 22 million pages.) It's becoming more and more complicated to find interesting information and the creation of the enhanced Web index proves it.

Is the Network Computer (NC) an answer to this? Take a look at your living room. What do we see? A TV, perhaps two, with one satellite decoder; your main telephone and perhaps your mobile with the battery for charging it; more and more probably your computer or your game station; and so on. Unfortunately you have no room for your network computer except if you convince your family to remove the guest armchair!

Coming back to business. In France, FT succeeded in placing a dumb terminal in each house because it was small and mainly free. If tomorrow you have to buy computer equipment and if you have the choice between a limited terminal and a powerful PC with multimedia extensions for £100 more, I think that 90% of you will choose the second solution.

The way for NCs to succeed is to be sponsored by telcos as FT did in 1980. It could be a good reason as the first beneficiary of NCs may be the telcos themselves due to the telecommunication consumption, but none of these telcos seems to be ready to reproduce large charity programmes, as in the past.

The latest trends seem to promise a great market for NCs inside the company because of the need for the majority of employees for only accessing the databases, filling in forms or producing documents. As the Internet mania turns into intranet, this is perhaps the right answer for the NC's future.

So, we need to wait for public equipment. This equipment must be cheap and have communication facilities (I mean not dialling with a 14,400 baud modem).

The first point: in June 1996 there was a press release from ACER for a cheap PC based on one-year-old technology. A lot of brilliant analysts laughed and claimed that it would fail. Is this not the solution? Does a family need a Pentium 250 MHz with 2 Gb disk? You can keep years of bank statements. The solution is perhaps in between an NC and a complete and modern PC: a lighter PC with enough disk space, a processor which is not the latest Intel version and a first generation CD-ROM.

Perhaps this is not a brilliant idea for computer manufacturers as they always want to sell the latest model, but if we take a look at the TV market we do not see this permanent technological race. We could also imagine that we will find some second-hand salesmen as we see for cars and this might not be a marginal market.

On the second point, communications, we'll have to wait for the generalisation of 'cable' plugs in every house. This is the only chance to combine high bandwidth and low price, when all home owners will have the same plug for TV and computer and will subscribe to its online service for less than £10 per month. In conclusion, if we want to create a home market, let's turn our mind to it and not to a technological market.

5. Think positive

So, everybody in France agrees on the fact that the business is in front of us. But, we already have a lot of lessons from this first period.

The first one is the rallying to a universal standard. The French are well known for their engineer mentality, wanting to invent all the technologies themselves. So now, we can forget the proprietary modems, the FT agreement, the 'not invented here' concept. We're in the world.

The second is what we can call the 'public service of the Internet'. French IPs were used to sell information, even marketing and sales information and even when it is from a governmental or public service organisation.

Will you accept to pay per view for advertisements on your TV? For an added value information like movies, yes, but not for marketing information like the price of your next car. After a golden period, people have the feeling that the Minitel was too expensive and the Internet arrives to rush into this. So we saw last year a lot of new services emerging on the Internet. They can be listed in three categories:

- IPs with added value information who are waiting for a payment solution;
- IPs with general information (public advertisements, movies, agenda ...);
- public service IPs.

The last kind have created free services (city information, tourist services ...) and the second kind have turned their mind to a new economical model where they sell Internet advertisements. This incidentally forces them to

build faultless services for attracting users like the TV channels.
And the winner is the user.

6. Conclusion

In today's business globalisation, will the French touch succeed in exporting its know-how, and what are the requirements for foreign companies to enter into this 10 billion Franc market?

On the first point, I think that the cultural exception is one key point for an online success, so if we exclude the large enough companies, wherever they come from (your side or our side of the Channel), to consider each national market as an entire and independent market, it will be difficult for French IPs to export their know-how (they are shy exporters) and for foreign companies to come in (the French cultural exception). Most opportunities will probably be on the IT side because of the capabilities of French companies to create and sell software solutions for online services, since they have ten years' experience in running services, providing tools for information managers, designing user interfaces ... For you, your chance will be advantageous to enter this large established market, taking into account our special nature.

Philippe Gilbert
Sales Manager, Computer Answer Line
92 Cours Vitton
69006 Lyon
France
Tel: +33 4 72831000
Fax: +33 4 72740485
pgilbert@cal.fr
<http://www.cal.fr>



Some features of the distribution of professional databases on CD-ROM in the Russian information market

Nickolay Romanov

Association of CD-ROM Users & Producers, Russia

Abstract: In spite of some problems and difficulties, the Russian information market of professional CD-ROM products is developing actively. There is potentially a large capacity for the Russian information market. This paper describes and analyses practical experience in distributing professional CD-ROM databases in the Russian market, and some its characteristic features. The features are determined by modern Russian legislation, the rules of financial regulation and the singularity of the economic situation. In particular there are the following features: specificity of current information requirements of the end Russian consumers; specific requirements of the distributors of CD-ROM products; the substandard forms of payment for CD-ROM titles; and some other features. The analysis conducted has produced some practical recommendations which should be taken into account in the planning of commercial policy. The account of these features, in our opinion, will allow the producers and vendors of CD-ROM products to sell professional CD-ROM databases in the Russian market more successfully

Keywords: Keywords: databases on CD-ROM, Russian market

1. Introduction

The features of the Russian information market, primarily the lack of reliable telecommunication networks, determine the long-term requirements for professional CD-ROM databases (Refs 1-2). Russian end-users of CD-ROM have a good technical basis for processing professional databases. Most large and average size state, municipal and commercial organisations have the means of playing CD-ROM in the form of 4-x or 6-x high-speed CD-ROM players from leading firms. Recently CD-NET servers have also begun to be distributed in the computer networks of Russian organisations.

The spectrum of information needs of the Russian consumer of professional CD-ROM databases is growing. Taking into account the potentially large capacity of the Russian information market and the increase of sales volume of CD-ROM databases produced by leading information firms it is possible to predict the future (Refs 3-4).

2. Main consumers of professional CD-ROM products and the sphere of their information interests

In Russia there is presently a wide spectrum of consumers and potential consumers of professional CD-ROM products. Analysis has shown that among the main consumers of CD-ROM titles are:

- state institutions (governmental and legislative bodies);
- information and scientific centres of ministries and departments;
- municipal bodies and establishments of large cities;
- administrations of free economic zones;
- universities;
- large libraries;
- large private concerns that provide information services and consultations.

The professional orientation of the main consumers defines the sphere of their information interests in relation to CD-ROM products. At present Russian consumers are expressing the most interest in CD-ROM databases in the following fields:

- economy and business (primarily, registers on firms);

- politics and law (abstract and full-text electronic titles, containing information from leading world periodicals, both newspapers and journals);
- science, engineering, medicine (abstract and bibliographical databases on world scientific titles);
- computer technologies (with reviews of the hardware and software market);
- help and encyclopedia titles.

3. Main criteria for the evaluation of CD-ROM titles by professional Russian consumers

The question of the evaluation of CD-ROM databases is now becoming significant. The professional approach to evaluation and choosing the necessary titles on CD-ROM has become widespread: chaotic and non-professional approaches are rare. The process of evaluating CD-ROM titles by professional Russian consumers involves a number of stages.

First of all, plenty of literature is analysed: for example, reference titles such as *CD-ROM Directory*, *Gale Directory* and so on; periodic titles such as *CD-ROM Professional*; and conference materials such as the Proceedings of the International Online Information Meeting. Then additional information is requested from firms: promotional materials, catalogues and so on. If the firm can send its title on CD-ROM for a 30-day free trial, so much the better. Such a form of interaction, in the opinion of Russian consumers, is the most effective. The main criteria applied by professional Russian consumers to evaluating CD-ROM titles are the following:

- *quality and accuracy of the information.* Unfortunately there are cases where CD-ROM databases contain inexact or out-of-date information;
- *completeness of the information.* Preference is given to the CD-ROM databases which contain worldwide information;
- *frequency of updating the information.* Many Russian consumers prefer information to be updated every month;
- *cost of the CD-ROM databases.* The market is analysed seriously and competitive prices of various firms are evaluated.

In short, all parameters are optimised and a balanced decision on the purchase of the CD-ROM title is accepted.

4. The official distributors

The correct choice of distributor for CD-ROM products is significant. In principle a foreign firm (producer or vendor of CD-ROM titles) can be a distributor: however, analysis has shown that Russian end-users of foreign CD-ROM titles prefer to subscribe through specialised Russian distributors. This is the most convenient from the point of view of the potential Russian consumers because of Russian legislation and the conditions of economic mutual relations.

This way allows end-users to interact effectively and quickly with the distributor in matters of delivery dates, price coordination and other questions concerning the subscription. Issues concerning the warranties of deliveries according to Russian legislation, and in Russian judicial and arbitration instances, can be settled. The Russian distributor can frequently carry out faster tutorial, technical and information support for the end-users. As a whole the Russian distributor is frequently more convenient for Russian end-users.

The Russian end-user can also pay the Russian distributor for the subscription in Russian roubles: it becomes the distributor's problem to convert the Russian roubles into dollars, pounds or any other necessary currency and to pay the producer or vendor of the CD-ROM products. With the background of Russian legislation, currency regulation and conditions of economic mutual relations this is frequently significant for the end-user.

The problem of fulfilment by the end-user of license restrictions for the use of the CD-ROM products is significant. In our opinion, the Russian distributor can supervise questions of the licensed use of CD-ROM titles more effectively.

5. Payment for the CD-ROM titles to the producers and vendors

There are a number of features in paying foreign producers and vendors for a subscription to a CD-ROM title.

At present in Russia payment for deliveries of foreign products using credit cards is not widespread. The prevailing system is for payment in the form of bank transfers. This situation concerns small-sized sums. It is particularly characteristic of state and municipal organisations, and also for many business concerns.

This feature of the Russian market is necessary for invoicing for the CD-ROM databases or for sending out

the subscription forms. It is necessary for the producers and vendors of the CD-ROM products to be guided by bank transfers, so full bank details by which payment should be made should be specified on the invoices and the subscription forms.

The producers and vendors also have to pay attention to a current condition of Russian customs legislation and currency regulation: the payer for a subscription for an electronic title on CD-ROM can be only that firm or organisation named on the pro forma invoice. This frequently creates additional difficulties for the dealers of CD-ROM products.

Producers and vendors of CD-ROM titles should also take into account such features of the Russian market as the instability of financial conditions and the risk of quick insolvency faced by potential consumers of CD-ROM products. This can often happen to commercial companies in Russia, for example because of changes in tax and regulating legislation, and also through losses from commercial activity, bankruptcy and other reasons. The solvency of Russian state and municipal organisations can also change quickly. If the financial assets of such organisations are not spent in the determined term (for example, by the end of a fiscal year or quarter), they can be withdrawn in the state budget. This is connected to existing legislation regulating the activity of Russian state organisations. Therefore money which was originally intended for subscription to a CD-ROM title may be spent on other needs. This is also characteristic for training establishments and libraries.

Therefore, it is necessary to recommend that CD-ROM producers and vendors answer subscription requests from Russian consumers and distributors as quickly as possible. It is necessary to send the requested invoices and subscription forms promptly; on the other hand, it is recommended to set the payment terms at no less than 90 days because of possible delays in payment. This meets with the real conditions of the modern Russian economy, and the situation with the tax and regulating legislation.

6. Language barrier

One of the specific problems facing a wide circulation of foreign professional CD-ROM databases in Russia is the language barrier. Many Russian experts (potential end-users of CD-ROM) do not speak English and other foreign languages perfectly, so preference is given to CD-ROM databases in Russian. This is especially the situation in the field of economics and business. In science and training the situation is much more favourable: the number of scientific experts who speak English is much greater.

Hence firms now delivering CD-ROM products in Russian will have significant competitive advantages in the Russian market. A positive example here is the policy of firms which develop software: for example, in the Russian market there is a great demand for Russian versions of Windows, Microsoft Word and other software products.

Nevertheless, it is possible to see that the Russian situation in the sphere of studying the English language has recently begun to improve considerably. The expansion of international contacts has resulted in a burst of interest in studying English. Therefore in the long-term the demand in the Russian market for CD-ROM databases in English will grow.

8. Delivery of CD-ROM products

In the Russian market significance is now being given to the discipline and method of delivery of CD-ROM products. End-users are most interested in faster and reliable reception of CD-ROM titles for which they have paid. Delays in delivery have a negative influence on the likely renewal of subscriptions.

The delivery of CD-ROM titles must be by fast and reliable mail. Russian end-users may frequently accept additional costs if they receive the information in good time. This situation is characteristic for CD-ROM databases in dynamically developing fields, in particular in the field of computer technologies.

7. Conclusion

Commercial distribution of professional CD-ROM databases in the Russian market is a complex process. For a successful commercial policy in the distribution of professional CD-ROM databases in the Russian market, producers and vendors must take into account the specific requests of the end-users of CD-ROM products. These requests are shaped by features of Russian legislation and the economic situation. Experience of distribution of CD-ROM databases in the Russian information market has shown an opportunity for an increase in the overall performance of CD-ROM sales if these features are taken into account. The analysis has shown sufficient serviceability and efficiency of these recommendations. Despite its problems, the Russian market for CD-ROM has good prospects.

Nickolay Romanov
Financial Director
Association of CD-ROM Users & Producers
8/2 Mjasnitskaja Street
101000 Moscow
Russia
Fax: + 7 095 9235549
E-mail: nikolay@cdromclub.msk.ru

References

- [1] Courage, M.-A. and A. Butrimenko (1995), Electronic information market in Eastern Europe 1995 — state of affairs and tendencies, in *Proceedings of the 19th International Online Information Meeting*, 5–7 December 1995, London, Learned Information Europe Ltd, pp. 297–303.
- [2] Romanov, N. (1994), Current state of affairs and perspectives of CD-ROM application in Russian research and development centres, in *Proceedings of the 18th International Online Information Meeting*, 6–8 December 1994, London, Learned Information Europe Ltd, pp. 529–533.
- [3] Romanov, N. and S. Saulyak (1995) Some CD-ROM projects in the Asian region of Russia, in *Proceedings of the 1st Asian Information Meeting*, 27–30 September 1995, Hong Kong, Learned Information Europe Ltd, pp. 105–108.
- [4] Romanov, N. and S. Saulyak (1995) Topical and prospective projects of the Russian Association of CD-ROM Users and Producers, in *Proceedings of the 19th International Online Information Meeting*, 5–7 December 1995, London, Learned Information Europe Ltd, pp. 365–369.



Development of online database services in Japan and perspectives on Asia

Takayasu Miyakawa

Mitsubishi Research Institute Inc., Japan

Abstract: The author outlines briefly the market developments, government promotion policies and efforts by private industries for online database services in Japan since the latter half of the 1970s. With these efforts and effects combined, the online database service market reached an order of US\$20 billion annually. Approximately one third of this market is Western online services. Then growth levelled off, partly because of the general economic stagnation since the early 1990s and partly because of rapid and profound change in the environment for information usage. New opportunities are provided by downsizing in computer usage, rapid progress in multimedia and digital technology, wide-spread information networks (especially the Internet) and a steadily emerging personal and home market. Experimental programmes for digital libraries have already started. The idea of end-user searches of digital libraries through a network is exciting but threatening for everyone in the online services industry. The author describes the current situation in Japan. In newly industrialising countries and areas in Asia such as Korea, Taiwan, Singapore, Thailand and Malaysia, online database producers and vendors are now emerging. The market has now started to grow, so within a decade there will be a fourth large market for online services in Asia after the American, European and Japanese markets.

1. Twenty years of growth of the online market in Japan

1.1. Initial phase (1970–1985)

The database service industry is a sub-sector of the information service industry which had annual sales of 6.2 trillion Yen (US\$62 billion) in 1994. Nearly 60% of the industry's sales are for software developments and programming: database services is the smallest segment. It has a history of over a quarter of a century and the domestic market has already reached 200 billion Yen (approximately US\$2 billion) in 1994 (Figures 1 and 2). The scale of the market is still not so large as an industrial sector, but it has been growing steadily in the past decades.

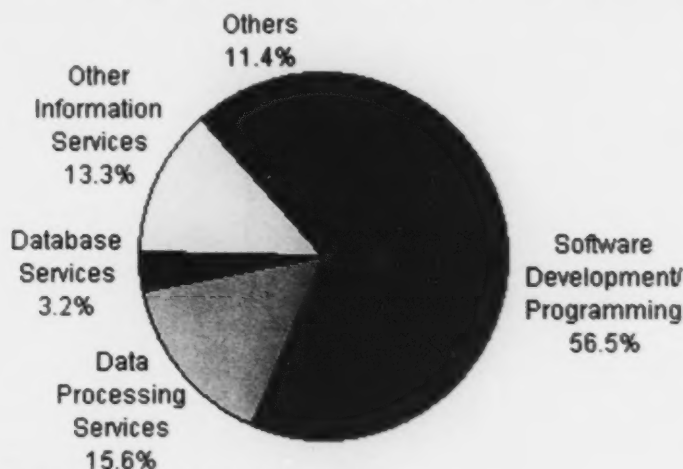


Figure 1: Sectors of the information service industry, 1994.

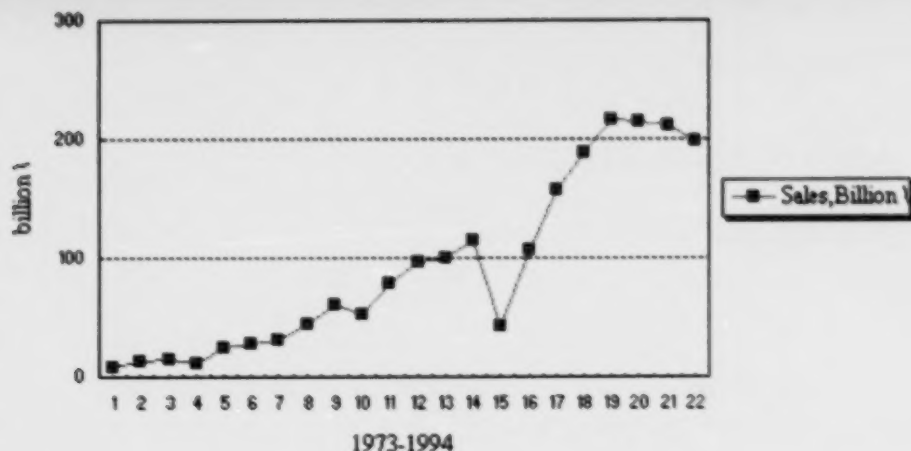


Figure 2: Database sales in Japan (billion Yen), 1973-1994.

There are three phases in the development of the industry. The first commercial database services were introduced into Japan in the early 1970s by several public organisations and private firms: for example science and technology information were introduced by JICST (Japan Information Center for Science and Technology); patent information by JAPATIC (Japan Patent Information Centre); and stock market quotations and companies financial data by Nihon Keizai Shimbun (Nikkei) and Quick. Foreign database services such as DIALOG were introduced to the Japanese market through representative agencies, Maruzen and Kinokuniya — the largest book retailers — in the first half of the 1970s. The technology upon which database services were based was a combination of large-scale host computers at the distributor's service centre and dedicated user's terminals connected with the hosts through low-speed government regulated public telephone circuits.

Government ministries and agencies created a set of promotion policies to make the electronic database service industry a vital part of the information infrastructure. One of the visible result of MITI's basic promotion policy was the establishment of the Database Promotion Centre (DPC), a public foundation devoted to supporting basic research and development of database production efforts financially, assisting regional promotion and training programmes, and so on.

1.2. The second phase (1985-1990)

Database services entered the second stage, the take-off phase, after 1985. In this phase there were a substantial number of new entries and several government promotion policies were implemented. According to the Database Directory (MITI), the number of database service providers (producers and distributors) exceeded 100 in 1985 and 200 in 1989. The number of databases commercially available to third parties in Japan exceeded 1000 in 1985 (Figure 3).

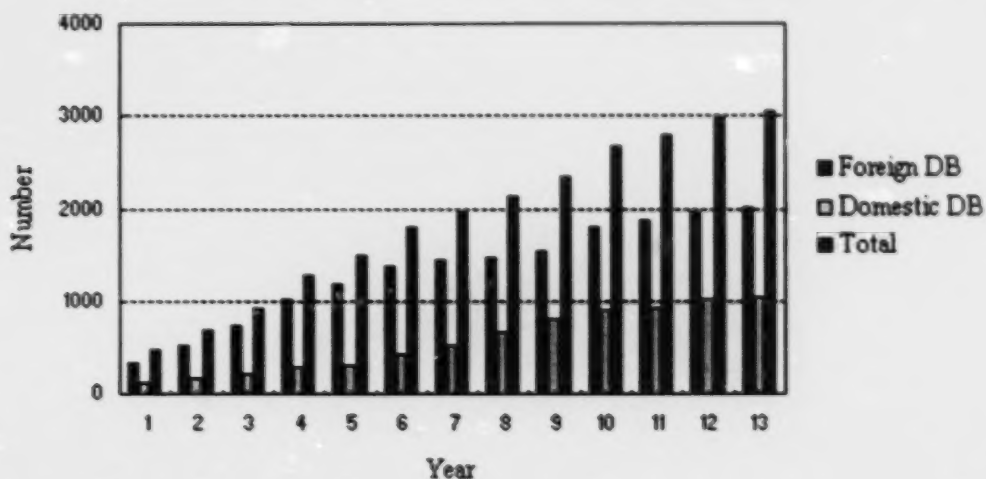


Figure 3: Databases commercially available in Japan, 1982-1994.

The most important policy relevant to electronic information services was the New Telecommunication Act of 1985 which privatised NTT (Nippon Telephone and Telegram Corporation, a state owned monopoly body) and started private VAN services. This was a trigger to the proliferation of new online services. It also stimulated private enterprises to construct their own in-house information networks.

In 1989, NTT's ISDN (Integrated Service Digital Network) began serving the public. In the area of academic information, the National Centre for Scientific Information System (NACSIS) was established by the Ministry of Education in the same year. NACSIS operates as a centre for a national network linking nearly 400 academic libraries — the first bibliographic utilities in Japan.

The technology basis of this second phase was the downsizing of hosts and terminals with the privatisation of telecommunication networks. Most customers of electronic database services were still in industry, so the databases were mainly concentrated in the areas of business and science and technology.

1.3. The third phase (since 1990)

For the third phase, since 1990, two things have had serious impact: economic depression and a fundamental change in the usage milieu. The Japanese economy experienced a relatively long stagnation after the collapse of the 'Bubble Economy' at the end of the 1980s, and the annual growth rate of the general economy in terms of GDP fell down to zero in 1993. These general economic trends were an unfavourable background for the growth of a market for information service industries.

Coincidentally, the milieu of the utilisation of electronic information services is changing rapidly and profoundly. Change has taken place in both the technology and the market. The conventional mode of use of commercial databases had begun to stagnate, and annual sales of database services — mainly online databases — had stopped growing at the level of annual sales of 200 billion Yen (US\$20 billion) (Table 1). Does this mean database services in Japan are saturated at the market limit? The author interprets that the structure of the market and the players within the industry are changing so rapidly that annual government statistics do not reflect these changes properly. Survey samples selected by the government are not appropriate for reflecting a rapid change of market structure, nor for including entries of new players. There is contradictory data and evidence that proves my supposition.

Table 1: Market growth in the early 1990s, Japan.

	<i>Information service</i>	<i>Annual growth %</i>	<i>Database service</i>	<i>Annual growth %</i>
1989	4351.4	32	157.6	3.6
1990	5826.7	35	188.6	19.7
1991	6886.1	17.3	216	14.5
1992	7127.6	1	214.1	-0.9
1993	6514.4	-8.6	211.5	-1.2
1994	6177	-5.2	198.8	-6

- (1) According to the user survey of the Database Promotion Centre in September 1995, annual expenditure per user firm or organisation for database usage in 1994 increased by 17% (large firm 16.3%, small and medium size firm 17.9%) on the previous year.
- (2) The providers' survey of September 1995 shows that even though 13% of firms replying recorded a loss, the average increase rate of sales of these firms was 25.4%. There is no news of serious withdrawal from the market among domestic database providers.
- (3) Despite the stagnant general economic condition, the number of databases available commercially is increasing steadily (Figure 3).
- (4) Access to commercial databases via PC networks is increasing. Frequency of use of database menus on PC networks increased from 35.5% in 1987 to 56.8% in 1994. The number of ID holders of PC network services reached over 3.7 million in June 1995. Therefore, the author's supposition is that due to downsizing technology and networking, the database service industry is now entering into a new phase of development after 1995.

2. Saturation and renovation since the early 1990s

2.1. Factors of change

Since the early 1990s the growth of the conventional database service market has levelled off, partly because of economic depression and partly because of rapid and profound change in the information use environment.

Multi-aspect innovation in digital technology, downsizing of computer usage, wide-spreading of information networks, construction of a high-speed digital circuit infrastructure, and a steadily emerging personal and home market are all factors which provide new opportunities. The idea of end-user searches at a digital library through a network is exciting but threatening for everyone in the conventional online services industries.

- (1) *Personal computers.* The strongest factor is the increasing availability of small, inexpensive but high-quality personal computers and workstations. With the recovery of the general economic condition, domestic shipment of PCs in Japan exceeded 5.7 million units in 1995. This is a firm sign of the digital revolution proceeding in every aspect of social and economic activities.
- (2) *PC communication networks.* Downsizing trends have drastic impacts on the mode of use of computers among end-users. The spreading personal computer network in Japan, represented by major network communication service suppliers like NIFTY-Serve and NEC PC-VAN, provides an effective gateway function to online databases. The 1994 survey shows that nearly three-quarters of providers expect database access via PC network to be the most common method five years from now. The next most effective new medium for database services is offline services by CD-ROM, followed by current online database services which are recognised by 40% of database service providers as the most effective way of accessing their services (Figure 4).

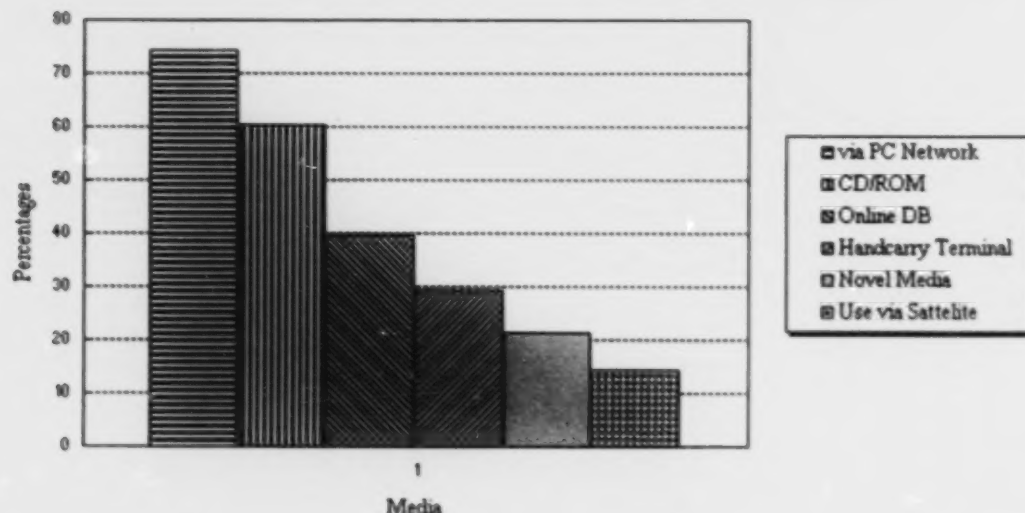
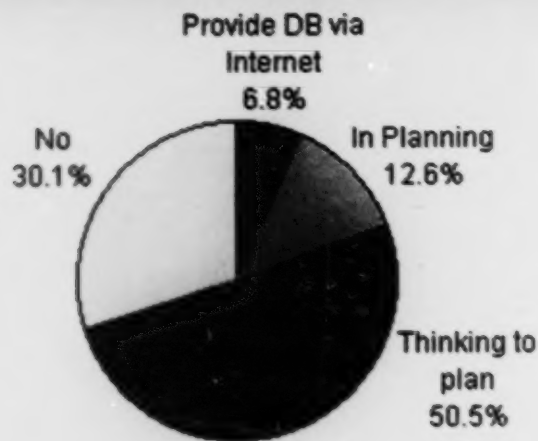


Figure 4: Expected means of database use in five years: providers' survey, 1994.

- (3) *In-house databases.* The downsizing trend makes it possible for users to build small and flexible local information networks of various types. These networks, in many cases under contract, allow end-users to access database hosts and download information through their PCs for their own purposes. Within the local area networks of private companies or public organisations in-house databases are compiled, maintained and used by employees. In-house database use differs by industry. Nearly 70% of private enterprises operate their own in-house databases and a significant number of them can be accessed by every terminal connected to the in-house networks.
- (4) *Internet.* Until the early 1990s, the Internet was mainly used in academic circles and several computer manufacturing firms as the vehicle for information exchange in scientific research. In 1992 the first commercial provider for Internet connection began to serve general users. As far as database services are concerned, the Internet provides new ways of accessing database resources with inexpensive access costs. From the user's point of view it opens vast access to commercial databases. Database providers find huge numbers of potential customers through the Internet. While database providers who allow users access via the Internet are gradually increasing, there are still only a few, i.e. 13% in 1995. If one includes those who are considering doing so it will be about 40% of the total database providers, but still the majority of providers express a certain degree of concern about the opportunities the Internet offers their services. The most frequent apprehensions they express are about security of communication, followed by the difficulties of fee collection, and ambiguous accountability when errors occur (Figures 5 and 6).

1993



1994

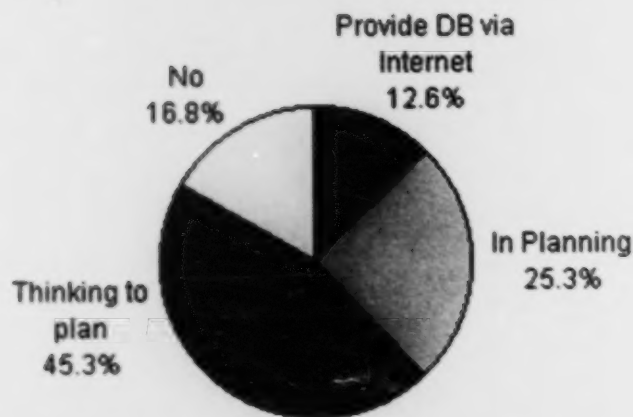


Figure 5: Database services via Internet, providers' view, 1993 and 1994.

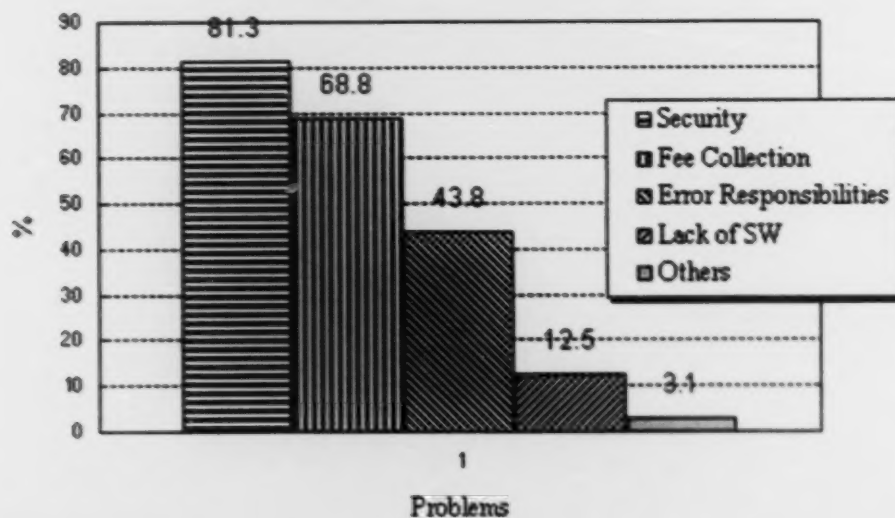


Figure 6: Problems of database services via the Internet, 1994.

BEST COPY AVAILABLE

Online Information 96 Proceedings

Page 335

- (5) **CD-ROM.** Some types of databases are converted to CD-ROM. These CD-ROMs have been increasing year by year since the early 1990s. The survey shows that providers of databases on CD-ROM increased from 20% in 1990 to over 40% in 1995. Those CD-ROM databases are mainly dictionaries, encyclopedias, directories, maps, catalogues and various indexes. Some 430 CD-ROM database titles have been made available to the public and an additional 130 new titles are being planned.
- (6) **Personal and home market.** Personal and home use of databases has increased in recent years through the spread of inexpensive and high-quality PC terminals. The Survey on Personal and Home Use in November 1995 reveals that 41% of usage is simply for personal purposes: 51% of respondents said this is for daily work. Home use opens a new market in which the files most frequently accessed are for searching newspapers and books. Favourites for young people are CD/Video Soft, Quiz/Game/Divination, Theatre/Movie and Education/Culture; housewives like Shopping/Seat Reservation, Cuisine and other Household-help files (Figure 7).

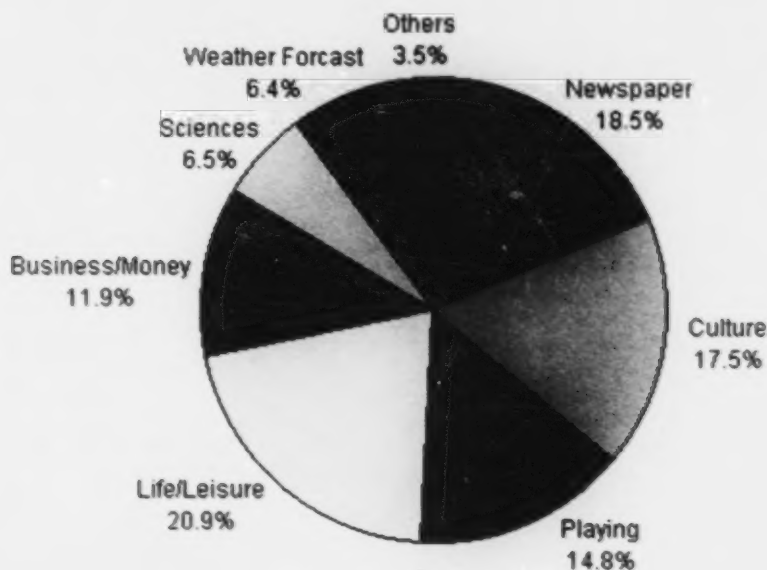


Figure 7: Personal home use of databases, Japan, 1994.

- (7) **Pilot Programme of Digital Library.** Experiments in digital libraries have already begun as pilot programmes by several public organisations, and the programme for a National Digital Library was announced in early 1995 as the Second National Diet Library or NDL-Kansai, for which the first phase of construction work has already started at the site in the Kansai Science/Culture City south of Kyoto. It is scheduled to open to the public in 2002.
- (8) **Kobe Earthquake.** On 17 January 1995 the city of Kobe was hit by a serious earthquake directly underneath. In a catastrophe people want information as well as food and beverages, and they realised the usefulness of the quick delivery of simple messages on several surviving PC networks. The role of personal PC networks in natural disasters or large scale accidents will be a new issue for crisis management in our society. The people of Kobe proved it.

3. New trends in Asian countries

3.1. A conceptual image of the stages of development for database usage in developing countries

Information technology has a tremendous impact on every aspect of economic and social activities in the newly industrialised and industrialising countries in Asia. Observing these developments from Japan, the present author dares to draw a conceptual image of the development phases of database usage. The first stage is a preparation period for the application of information technology in economic and social activities. The government designs a promotion policy for information utilisation. Pioneering efforts for database use and local data file compilation are put forward, mainly in the public sector, for example by government organisations and universities. A high degree of dependency on imported Western database services continues for a decade or more.

In the second stage, government efforts are concentrated in building a domestic information infrastructure, extending telecommunication networks, and promoting and training the professional capabilities of information specialists. Database producers still lead in the public sector, such as National Science and Technology Information Centres, and related information dissemination systems such as administrative, academic and patent information. In the private sector, database services usually start in the areas of stock exchange quotations and newspapers.

In the third stage of development of database services, private firms enter the industry and the production of domestic databases based on vernacular languages becomes a major part of their effort. The online market for electronic databases in various subject areas is supported by widespread use of PC terminals among administration, academies and business enterprises. The information industry, including computer hardware manufacturing, is firmly established as a new industry in its industrial structures. The author observes that three countries and economies in East Asia — Korea, Taiwan and Singapore — have already entered the third stage of development in recent years.

3.2. Korea

3.2.1. Production of Korean databases

The production of electronic databases in Korea began in the early 1980s, and the number of databases produced and distributed commercially to third parties in Korea reached 1061 titles as of the end of August 1995 (Directory of Korean Databases, issued by the Korea Database Promotion Centre). Compared to the previous year, 1994, in which the number of database titles available was 907, the rate of increase is over 30% within a year. These databases are produced by 438 various organisations of which government organisations account for one fifth, private companies one fourth and specialised information service companies one sixth. In recent years, increasing numbers of organisations and institutions in the private sector are entering into information service industries. Sales of commercial databases in Korea reached 74.6 billion Won (approximately US\$96.3 million) in 1995, including sales of Korean produced databases and those for imported databases as well (Figure 8). Due to stagnant general economic conditions in 1993 and 1994, sales of database services fell. Perhaps this could be interpreted partly as the impacts of downsizing and networking technology on domestic information services.

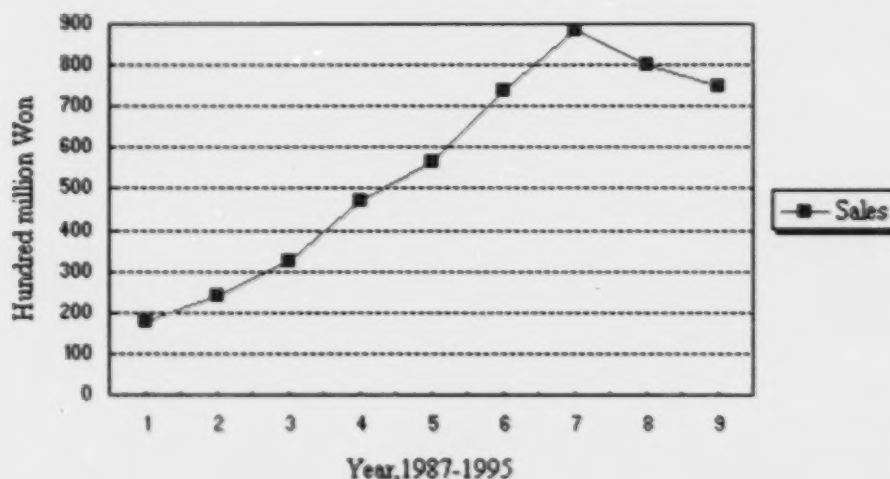


Figure 8: Sales of databases, Korea, 1987-1995.

The number of databases relating to industrial and business information has been increasing since the early 1990s, and reached nearly 40% of the total number of 1061 in 1995. Those for social life account for one fourth, i.e. over 100 databases. Seventy three cover science and technology, including medical science, i.e. 6.9% of the total number in 1995 (Figure 9).

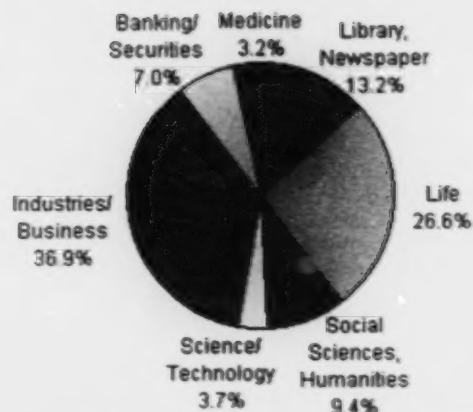
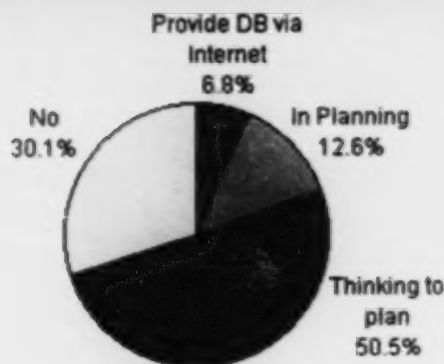


Figure 9: Subject areas of Korean databases, 1992 and 1995.

3.2.2. Distribution systems in Korea

There are 51 organisations and companies distributing online databases, of which 13 major vendors provide online services. A substantial part of commercial services, including those for Western and Japanese databases, are concentrated on large telecommunication companies such as Datacommunication Company of Korea, and specialised database services firms like Korea PC Communication. DACOM's service system is called 'Chonli-Won' (meaning Clairvoyance), which mounts 396 files, and Korea PC Communication's 'Hitel' serves 391 files. Then come service systems such as NAUCOM's NAUNULI, Atel, Lucky-Goldstar's GINS and KINITI's (Korea Institute for Industrial and Technological Information) KINITI-IR.

3.3. Taiwan

3.3.1. Production and distribution

Taiwan is an important area of China where foreign reserve exceeded over US\$80 billion in 1995 — one of the most economically prosperous areas in Asia. Information service industries have recorded remarkable growth since the early 1990s, which reflects the growing manufacturing activities of electronic products and peripheral parts. Online database services began in Taiwan in the early 1980s: domestic producers' and vendors' host systems increased gradually, and database files available to the public exceeded 500 titles in 1994. Annual sales of these commercial database services are reported as 1500 million New Taiwan dollars (approximately US\$50 million) in 1995 (Figure 10).

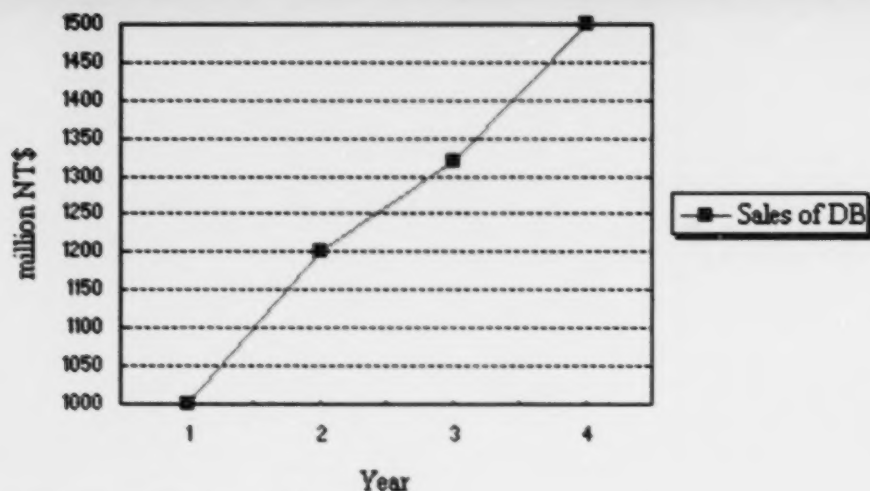


Figure 10: Sales of databases, Taiwan, 1992-1995.

In 1994 there were 180 public institutions and private companies producing Taiwan domestic databases, and 56 distributors serving domestic, Western and Japanese databases.

The largest share for subject area is over 40% for business, which covers finance, economy, market trends, credit information and foreign trade (Figure 11). This reflects the active business of Taiwan as the most vital trading area near the Chinese mainland and the Japanese archipelago. Commercial firms tend to pursue and collect basic economic and financial data of foreign importing corporations and wholesalers. Developments and trends of import/export regulations of foreign governments, patents and trade marks information are important to them. Academic institutions are one of the most intensive users of online service, therefore a quarter of available databases cover science/technology and medicine.

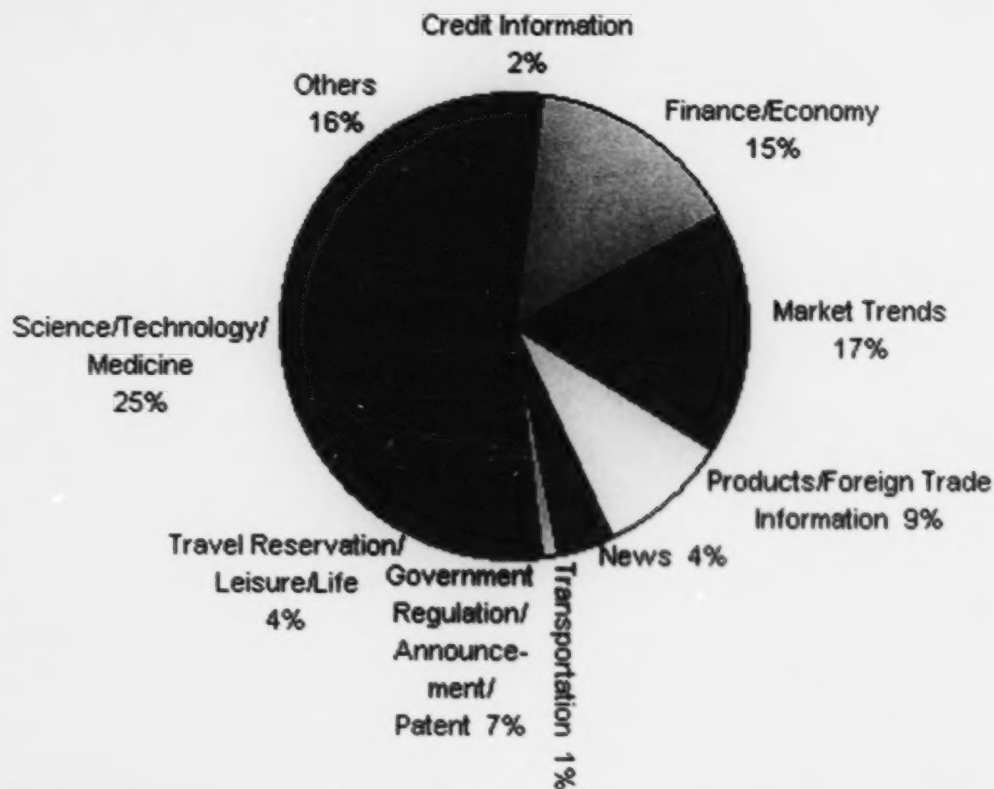


Figure 11: Database subjects area, Taiwan, 1994.

While the number of databases relating to consumers' life and leisure is still few, it is seen to be growing due to the improvement of living conditions and people's abundant disposable income.

3.3.2. Government's promotion policy

Since 1992 the Taiwan government has pursued the Five Year Plan for Promotion of Database Usage and Dissemination. Tasks for Taiwan database service providers are as follow. From the database producer's point of view, the most pressing task is to improve the quality of domestically produced databases and to enhance providers' technical capabilities, especially software technologies for improvement in user-interfaces.

As for tasks related to users, education in effective searches of deeper levels is the most important. Training professional searcher service consultants who can intermediate between database vendors and end-users is another target.

The most imminent task for the government related to the enhancement of the new information infrastructure is a gradual deregulation of telecommunication regulation. This is to prepare a new telecommunication infrastructure for a much higher level of networking, such as Internet connections. Revision of Copyright Law and new issues of regulation to protect personal data from unfair use of electronic data files are other jobs for the administrative agencies.

3.4. Singapore

3.4.1. Growing information industry

The information industry is one of the key industries in Singapore. Since the late 1980s the annual growth rate of sales has been nearly 30% (29.5% on average between 1987 and 1994). The information industry consists of three major groups: hardware manufacturing, software development and information services (Figure 12). Hardware manufacturing and assembling activities are predominant — over 70% of total sales. Being in a favourable location as the regional centre of the flow of trade and traffic, and with a heritage of a long established mixed cultural and social milieu of Malaysians, Indians, and Chinese, Singapore is becoming the modern economic centre of trade and information in this part of the world.

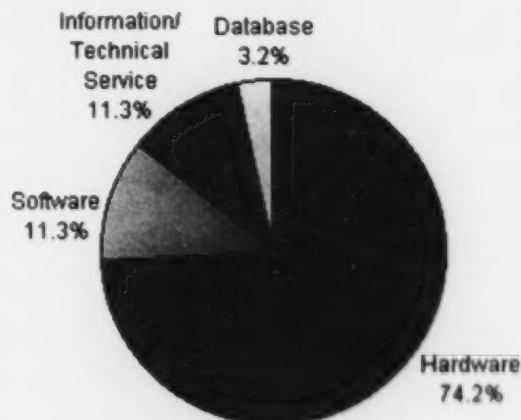


Figure 12: Singapore information industry sales, 1994.

The government information policy, IT 2000, aims to remodel Singapore as an information-oriented, high-tech equipped country in the coming century. The information industry's share of export to neighbouring countries such as Malaysia and Indonesia has been gradually increasing since the late 1980s, and now exceeds over 40% (Figure 13). Information technology services are about 15% of the information industry — over 750 million Singapore dollars (US\$490 million) — in 1994, which include various professional services such as out-sourcing facilities services, account auditing, technical training and consulting, and online database services.

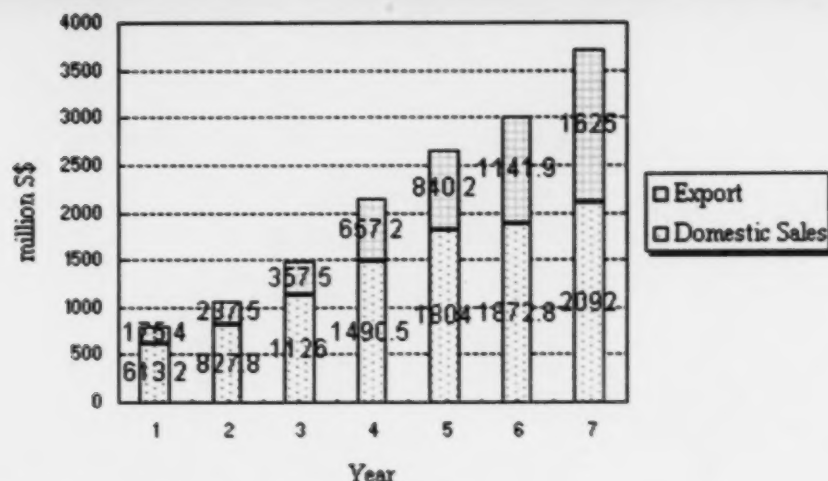


Figure 13: Information industry sales, Singapore, 1987-1994.

3.4.2. Database services

Sales for database services in 1994 were 21% of the sales of information technology services, i.e. 157.8 million Singapore dollars (US\$103 million). This include sales for foreign clientele.

Online database services started with the Singapore Libraries Catalog System in 1988. In the same year Singapore Telecom started its videotex service system, called Teleview. Next year in 1989, the VAN service of SNS began to operate its networking service, connecting various online database services — such as legal, medical and financial — on its Tradenet (EDI). SPH (Singapore Press Holding Co), a major database service provider, inaugurated a newspaper information online service called News Link in 1994. This now covers newspapers published in Singapore in English, Malay and Chinese, and others published in Singapore, Kuala Lumpur and Jakarta. Furthermore, SPH recently opened the 'Asia One' home page on the Internet which covers newspaper headings and articles, and Singapore stock market quotations, travel and leisure information, classrooms announcements and so on.

3.5. Thailand

Commemorating the 50th year of the reign of the King, the Thai government designated 1995 as The Year of Thai Information Technology. The National Information Technology Committee (NITC) drafted a basic policy on national information, complying with the Seventh Development Plan of Thailand. NITC has set three basic targets for the information era:

- (1) construction of a National Information Infrastructure. A nationwide telecommunication network will be constructed by the year 2000;
- (2) education and training of specialists for information technology;
- (3) enhancement and improvement of government services.

The Computer Association of Thailand (CAT) and its Vendor Group (CAT-VG) jointly carried out a comprehensive survey and estimation of the Thai information industry in April 1995. According to this survey, information industry sales — including hardware, software and information technology services — were estimated at 29,872 million Baht (approximately US\$1187 million) in 1995. The sales increased at a rate of 37% on the previous year, CAT reports. The size of the market for information technology services was estimated as 2336 million Baht (about US\$93 million) in 1994 and 3480 million Baht (about US\$139 million) in 1995. This is a significant increase of 49% annually. Since the late 1980s, government technological institutions and universities have been producing electronic databases and information networks, and organisations in the public sector have been leading efforts. Private database service firms have just started offering their services such as stock market and investment information, newspaper headings and industrial information from government agencies. Several Western vendors like Dow Jones News Retrieval and Reuters' s Business Information are active through their local agents' firms. The commercial online market has just opened in this country.

4. Concluding remarks

The market for conventional online services of databases in Japan, approximately US\$20 billion in annual sales, is undergoing much change. Giant new players such as broadcasting corporations and software developing

corporations are now entering into the borderless market of information services. Digital libraries, electronic newspapers and CD-ROM books could be accessed, at relatively low cost, through Inter- and intranet services by customers who operate their own personal computers at home and in the office as well. What, then, is the role of information professionals who intermediate a huge population of end-users with information resources?

On the other hand, newly industrialised countries in Asia are entering into the information age. The scale of market for information services is still small but it is growing rapidly (Figure 14). It will reach a substantial size as a self-sustaining industry. Twenty years ago, when online database services appeared to the public, nobody could have imagined the current situation.

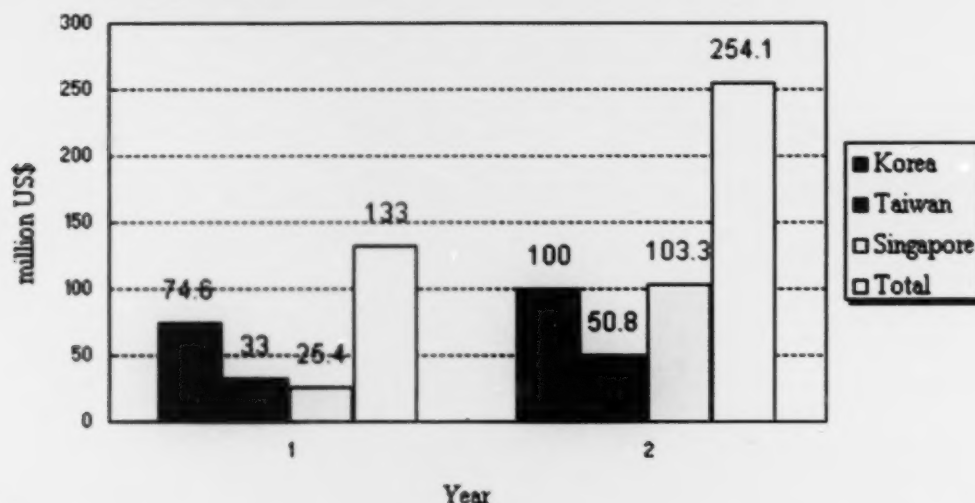


Figure 14: Sales of databases, East Asia, 1991 and 1994.

Takayasu Miyakawa
 Counselor, Mitsubishi Research Institute, Inc.
 2-3-6 Otemachi
 Chiyodaku
 Tokyo 100
 Japan.
 Tel: +81 3 3277 0847
 Fax: +81 3 3277 0884
 E-mail: MXH03272@niftyserve.or.jp

References

The author mainly relies on the information and data published in the Database Promotion Centre's Database White Paper, in Japanese, every year. It devotes one chapter to overviewing developments in Asian countries.

The latest issue is: Detabesu Shinko Senta, Detabesu Hakusho, (Database Promotion Center, Database White Paper), 1996, Tokyo, 369pp.



The end-user cometh and cometh again and again

David Nicholas

Information, Technology, and Change Analytical Research (ITCAR) Group¹, City University, UK

Ingrid Frossling

Department of Economics, University of Stockholm, Sweden

Abstract: *The bogey of the end-user has haunted the information profession since the appearance of the first online services in the early 1980s. During the 1980s the profession constantly debated the outcome of end-use — some doubting whether it would ever happen and others forecasting the Apocalypse. Since then the likes of Textline, NEXIS and FT Profile have been joined by waves of even more user-friendly CD-ROMs; and most recently we have been visited by the Internet. Nobody is questioning the permanence or fickleness of end-use anymore, but few seem to be debating its consequences. This article examines the impact of three rounds of end-users on the information profession and looks ahead to what the future holds.*

Keywords: end-users, online services, the Internet

1. Introduction

The Internet represents both a threat and challenge to all professional and mediating groups, but especially to those in the information industry — journalists, publishers, booksellers and information professionals — for it has the potential to turn their worlds upside down and inside out. Although it has to be said that it probably poses the greatest and gravest threat to information professionals and specialists. There is every chance that this group will become the first major road casualties along the information superhighway. But maybe information professionals have become complacent as a result of having beaten off the first and second end-user incursions on their territory: the first triggered off by the appearance of full-text online services like FT Profile, Textline and NEXIS, and the second by the arrival of the allegedly user-friendly CD-ROMs; for there appears to be very little end-user debate in the professional press these days. Or maybe it is simply a question of being resigned to their own fate? Using the experiences of those who were in the forefront of the first two waves of end-users, this article will examine the impact of the third, and possibly terminal, incursion — that occasioned by the Internet.

2. The end-user waves

2.1. The impact of the first wave: full-text, natural language online systems

What lessons can be learnt from the first wave of end-users that entered the online information domain in the mid-1980s on the back of full-text hosts like NEXIS, Textline and FT Profile? A number of pundits at the time felt that this wave would be terminal for the profession. Encouraged by the early successes of these full-text services with end-users, commentators began to write the epitaph of the intermediary. Jones, for instance, thought that 'the evolution of the professional online searcher must be regarded as a temporary aberration' (Ref 1). A quote from a *Financial Times* business report on the online industry gives a flavour of much of the writing at the time: 'The emphasis is expected to shift further away from the librarian and researcher to the end-user because of the current explosion of microcomputer installations and increased user friendly software' (Ref 2). History tells us that these comments were wide of the mark. In most cases what actually happened proved to be very different. Consider the experiences of newspaper libraries, the early leaders, who met head-on the challenge of a proficient group of information handlers being given direct access to a full text host with their own newspaper on it (FT Profile). Despite early anxieties, those newspaper librarians who responded to the threatened loss of their monopoly of the newspaper cutting archive by embracing end-use and not fighting it saw that the prime impact of end-use was to increase the demand on mediated searching (Ref 3). Through end-use journalists came to

understand online's potential and the librarian's skills in searching online databases. It turned out that the vast majority of end-users also wanted to delegate — usually the more difficult searches or those for hosts they were not familiar with. In broadcasting, where end-use had never gained anything more than a toe hold, libraries were devastated — many disappearing and the remaining ones shrinking to one or two person bands, usually left looking after the video collection (Ref 4).

Turning to the end-user experiences of other leading professional groups, the experience was much the same. MPs at the House of Commons, despite being given unprecedented access to online systems, wholly eschewed end-use, even when it concerned so-called user-friendly Textline and FT Profile (Ref 5). Shortages of time and a willingness to delegate meant that they were more than happy to get the Library, or their own research assistants, to search on their behalf. Even their research assistants, despite taking to online searching with some alacrity, would delegate the search to librarians in increasing volumes — a case of the online search being delegated twice. Take, too, the case of lawyers. LEXIS, the full text legal online system, came on the market with the expressed purpose of excluding the middleman (the librarian) from the legal information chain: in the beginning the company behind the service, Mead Data Central, would only provide the service to law practices and schools and would not train librarians in its use. They failed miserably in their attempt and now LEXIS wholeheartedly embraces librarians, probably their biggest customers now. Lawyers have shown little interest in searching databases themselves.

Some information professionals missed out completely on the first wave of end-users, and they were largely public librarians. And it was simply not just a case of lack of money that saw them duck the opportunity, for even when they were offered near-free access to an end-user online service that fitted their community profile — Volnet — they still would not take the plunge². It seemed that they were frightened to change, to innovate.

2.2. The impact of the second wave: CD-ROMs

It was generally thought that CD-ROM was a much more appropriate end-user tool. You could let end-users loose without fear of bankruptcy. It was also alleged that because it was such a user-friendly medium, what with its menus and windows, there were no costly training implications. More importantly, perhaps, it fitted into the existing scheme of paying for information — you paid up front, on an annual subscription basis, just like for a serial. In consequence many public libraries belatedly entered the fray. A delay, it will be argued later, that cost them dear.

For some end-users the advent of CD-ROM meant the loss of access to online systems. Many universities cut back on online and increased CD-ROM expenditure. Indeed, the great growth in CD-ROM use has largely come courtesy of university students. Other groups of user were not so easily pleased. Journalists almost wholly absented themselves from the technology, citing currency problems, poor retrospective coverage and limited range as the reasons. Politicians are sceptical. At the House of Commons, for instance Reuters-Textline was replaced with CD-ROMs that have since been networked via the Parliamentary Data Video Network (PDVN). A questionnaire study of pilot PDVN users — MPs and their research assistants — showed that they were not all that impressed, rating fax as the most useful feature of the PDVN: CD-ROMs came a poor third after POLIS (their own library OPAC) and they bemoaned CD-ROM's lack of currency (Ref 6).

This time around nobody disputed the end-user credentials of the technology — it was almost as though this was an article of faith. But despite the greatly increased access to computerised systems provided by CD-ROM, and its widely claimed user-friendly characteristics, there were still people — sometimes very eminent ones — who encountered problems with the technology. Take the published observations of one medical librarian: 'I have frequently found people searching [Medline CD-ROM] in the most peculiar, inefficient and totally illogical fashion. It is a slightly hair-raising prospect when the searcher is a junior doctor sent by his chief to do the spade work for an article to be submitted for publication, let alone searches relating to treatment of actual patients' (Ref 7). Nor was anybody talking of end-users taking away jobs anymore. It could be that information professionals had learnt from their experiences with online: CD-ROM was creating more work, too. More likely, the fact that this technology was not so powerful made them feel rather more secure: online was for the information professionals while CD-ROM with its poor currency, limited archive and coverage was very much a system for the end-users. A case of first class and second class information services. Now, that seemed to fit the natural order of things.

Of course, CD-ROM use is still growing very rapidly and it is only the arrival of the Internet that has put it into the shade. Indeed, there is a danger that a profession so fixated by the Internet will fail to take account of the rapid march of the CD-ROM. The figures speak for themselves: in 1990 817 titles were published; by 1995 the number had reached 9691; in 1992 unit sales reached 2 million; just two years later the figure was 27.8 million³. By any standards this is truly spectacular growth.

2.3. The impact of the third wave: the Internet

Just when CD-ROM looked to be burying online as an end-user tool along came the Internet. Interestingly, all the things that were apparently wrong with online — costly phone lines, fragile and complex modems etc. — were suddenly all the right things with the Internet, because through these things instant access to current data on a huge world-wide scale could be obtained. The Internet does herald a new end-user dawn. For a start, while the first wave saw thousands of end-users entering the marketplace and the second wave tens of thousands of users, with the Internet we are talking millions, if not tens of millions. End-use has jumped off the scale, and in a tiny fraction of the time it took online hosts to obtain their modest base. When millions of people start using online services for the first time, and in a very short space of time, it has to make a big wave. It would be naive in the

extreme if we did not think this was going to have a major impact on the information profession, for after all they are, arguably, standing closest to the wave. Undoubtedly, people will do a lot more of their own searching and it naturally follows that they will do some of the searching they once delegated to information professionals.

Irrespective of its scientific and academic beginnings, with the Internet you have an information system firmly rooted in the end-user domain. End-users clasp it firmly to their bosom: nobody mentions it as a tool for the intermediary, which would be politically incorrect. Even intermediaries seem certain of its end-user provenance. Some do throw bricks at it: lack of authority, poor response times, insufficient precision and lack of security are the ones thrown most often (Ref 8). But the bricks just bounce off, largely unnoticed by the adoring hordes of end-users and other information handlers, like journalists. Coming to it through Netscape, Lycos and so on you are constantly reminded of its end-user pedigree by the mouse-driven Windows environment, advertisements; colour, pictures, and simple-like-falling-off-a-log search systems. With the Internet comes information systems that reflect end-users' rough and ready searching styles, with much being left to browsing. But information professionals have to get away from rubbishing end-user searching styles and systems, and start training them to get the most from these systems.

For once, and for the very first time with an online service, the end-user and the intermediary are competing on a level-playing field. Both groups have unrestricted and unparalleled access to information. Training is no longer the key determinant here: luck, experience and time are probably among the key components of search success. The fact that the Internet, if not free, is cheap will also undoubtedly play an important role in the end-user will-they-won't-they debate. With cheap or seemingly free online access another potential barrier to end-user access has been demolished. As librarians were originally responsible — and still are in many cases — for providing access to documents then the fact that the Internet leads to direct document delivery will spell the real challenge: might information handling then become the gift of the individual worker?

The Internet also brings with it opportunities but only for those information professionals that grasp them quickly. In a world of rapid and constant change, information professionals cannot sit back and let others make the running as some suggest. If anything has been learnt after more than two decades of IT watching then it is that if you (1) wait until the technology stabilises, you will wait for ever; or (2) let others lead, then they will take all the spoils. Thus the Internet greatly raises the profile of online searching and it increases people's awareness of online retrieval (and their frailties and limitations in this regard). This can result — as in the case of online — in greater use of the information specialist, too. Furthermore, the Internet introduces a whole new army of users to the world of online — the information unaware. Through the Internet they will also be introduced to the world of intermediaries — a whole new market for their services. In addition, the Internet acts as a marketing and recruiting tool for the information profession. Increasingly prospective library school students, when asked the question of why they want to join the profession, reply that they want to learn about IT because that is where the future lies — and by IT they mean, online, CD-ROM, the Internet et al.

Challenges, but also changes: information professionals will have to become more specialist, more trainer, more proactive and more reference specialist: moving between the various information channels, balancing one up against another and choosing the most cost-effective option. The first wave of end-users gave information professionals time to adapt and change, and to ready themselves for the really big challenges to come. It gave information professionals time to show that they had expertise and authority in IT matters; it gave them time to create a new market for their services. Those who took the opportunity earlier are very well placed to handle the challenges of the Internet: their users will turn naturally to them for help in the times ahead, but for the others it is probably too late — they will be swept along by the tidal wave and maybe swept away altogether (and there will be casualties). Even amongst those who have profited from online and CD-ROM there is no room for complacency. This wave by contrast is of tidal-wave proportions — we have not seen anything like it yet.

Of course, increased accessibility to information is one thing but wanting to search or find information is something else. Before the Internet came along users were not exactly short of information — the fax, e-mail, online databases and the phone were already taxing people's ability to absorb information. Indeed, it could be argued that in the long run the Internet and its ilk will prove to be counterproductive: they will boost considerably the host of information want-nots. It is most likely that it will be those people deprived of these systems and communication forms (the home user) or those people for whom access to information gives them the edge or trade (business people and journalists, for instance) who will pitch in most actively.

There is a downside to the rapid and widespread acquaintance with end-use: lots of people become instant experts. The acquaintance with the Internet makes them an online expert: an information expert, even. This qualifies them to make judgements about all online purchases — CD-ROMs, commercial hosts etc. Here is where the authority and expertise of the information professional will be put on the line. At one newspaper, which will have to remain anonymous to protect the identity of the librarian, the Managing Editor is dictating online policies and purchases, against professional advice and wisdom, obtained as a result of ten year's experience with online. The end-user journalists will be the losers. One can see the same thing happening in regard to computer professionals, emboldened by the huge interest in computer networks. Already in the States library schools have closed as students have looked to computing for the IT skills needed in today's information society.

3. Public libraries — the biggest threat

Of all information professionals, public librarians face the greatest threat from the digital revolution. And they largely have themselves to blame for this. They largely rejected the advances of online in the 1980s, partly

because it was an affront to some of their principles, like charging for information, for instance; partly because it was seen to be elitist; and partly because it was new, and public libraries have never been very good at new product development. That meant they made a very late start. By not clambering aboard the online bandwagon earlier they lost their role and authority (and gained no reputation) for overseeing machine-readable data. A belated attempt to get involved in CD-ROM, a technology potentially as expensive as online but easier to administer and police, is also running into difficulty. They cannot make up their minds as to their role: lender, trainer, searcher or shopkeeper. Some indeed question whether they have any role at all. A hard-copy mentality and a hopeless mission to meet all the disparate needs of their users acts like a ball and chain on their efforts. Thus it is argued that, because the elderly are the key user group in most public libraries, there is no point introducing technology for this group would not be receptive to it. Even if this was true, the plain fact is that the elderly are such an important group because public libraries provide what they want — lots of hard-copy fiction. If you filled libraries with Internet terminals and CD stations then things would soon change, and you would find that the young would become the chief customer. But in a world without elastic budgets you have to let go of one group to get to the other, and that public libraries have manifestly been painfully slow to do. What is plain, though, is that the answer is not to de-professionalise the service, which appears to be the strategy of an increasing number of local authorities.

4. Re-defining the end-user

So many end-users have entered the marketplace that there needs to be a re-examination of the term 'end-user' itself. The minority are now the majority. In the very early days of online end-users, computer enthusiasts meant little more than a few keen scientists and academics. Then came the professional groups, lead by journalists and business people. These were followed by students in very large numbers, and finally by the general public, who are also likely to enter the fray in even larger numbers. Given such heterogeneity and the sheer size of the group — at one time very much smaller in number than the information professionals and now totally dwarfing them — it seems sensible to divide them into more meaningful and more homogeneous groups: academic end-users; practitioner end-users and the general public as end-users, for instance.

5. Conclusions

When the senior author of this paper first started researching end-users in the mid-1980s there were probably no more than a couple of hundred end-users in the UK; now there are probably more than a million. In the circumstances it is remarkable that the library (and information profession) as we know it has stood the passage of time. But one possible reason why the traditional model of a large, centralised physical collection supported by online searching professionals has stood the test of time so far is because there really was no genuine alternative (threat) until very, very recently. The real test will surely come in the next three years.

With the Internet there will undoubtedly be some losers and some winners, rather as was the case with online and CD-ROM. It will not herald the end of information professionalism but it will change its shape irrevocably. The Internet is capable of delivering a genuine and suitable information alternative. What will compensate for the loss of users to the Internet will be hordes of users who will enter the information environment for the very first time: they will need help, training, protection and advice (services not always traditionally associated with the librarian). There will be a need for someone who embraces and co-ordinates all the various information channels/media. There will be a need for someone to do the difficult searches, the searching for people who have simply not got the time to search — and there are lots and lots of them.

There will be increasingly a need to customise, tailor-make and validate information flows — there is too much information around for people to find their way around and to absorb it when they find it. With increased and unlimited access comes the inevitable problems of information overload, and the information anxiety that results. We have moved rapidly and frighteningly from under-provision to massive over-provision of information. With such large volumes of open-access data comes too the problems of inconsistent data, errors and useless or redundant information. Paradoxically the high tech solution — initially anyway — to a largely high tech problem seems to be the creation of faster machines, with more capacity, faster modems or dedicated lines and so on to reduce the time the user is searching, processing and waiting for the information. But the increased speed simply provides more information in a shorter time span. We are undoubtedly running up against our physiological limits in terms of the information we can process and absorb. As Toffler (Ref 9) notes, our capabilities in this respect are finite. The latest high tech solution is 'intelligent agents': anything it seems but the established and trusted alternative — the librarian, who will probably still be there after the 'intelligent agents' have long gone.

Finally, there will be all the added work that comes from end-users obtaining a greater awareness of the wealth of electronic information services that are out there. Security, standards and indexing concerns also have not gone away; indeed, if anything these concerns are greater now than ever before. The requirement though will be for more specialists, more computer-qualified people, working in smaller groups — and just probably for less of them.

Note

The authors are currently investigating the impact of the Internet on journalists and media librarians. The oral presentation will include some findings from this research, to support and illustrate the points made in the article.

David Nicholas

Senior Lecturer, The Information, Technology, and Change Analytical Research (ITCAR) Group

Department of Information Science

City University

London

UK

Footnotes

- (1) The group is multi-disciplinary and international in character and contains both academics and practitioners. It also counts among its membership Professor Peter Cole, School of Journalism, University of Central Lancashire; Helen Martin, Chief Librarian at *The Guardian*; Colette Batterbee, Department of Information Science, City University; and Tom Dobrowolski, Institute of Library and Information Science, University of Warsaw.
- (2) Volnet hosts a number of databases from community and voluntary organisations, many of direct relevance to the work of public libraries. It was offered to public librarians in its online form for less than £100 a year, unlimited use. There were few takers.
- (3) Source: TFPL (1993-95).

References

- [1] Jones, K. (1984) The effects of expert and allied systems on information handling: some scenarios, *Aslib Proceedings*, **36**, 217.
- [2] Holmes, N. (1985) *Online Information in Europe and the US*, Financial Times, London.
- [3] Martin, H. and D. Nicholas. (1993) End-users coming of age? Six years of end-user searching at *The Guardian*, *Online & CDROM Review*, **17**(2), 83-90.
- [4] Nicholas, D. and P. Pandit (1994) What happened to libraries in Independent Television? *New Library World*, **95**(1114), 4-7.
- [5] House of Commons Information Committee (1994) *The Provision of a Parliamentary Data and Video Network*, HMSO, London.
- [6] Nicholas, D. (1995) *An Assessment of Stereotypical Models of Online Searching Behaviour: Practitioner End-Users*, City University, Department of Information Science, London (PhD thesis).
- [7] Brewster, P. (1995) *Letter Information World Review CD-ROM*, May.
- [8] Kelly, S. and D. Nicholas (1996) Is the business cybrarian a reality? Internet use in business libraries, *Aslib Proceedings*, **48**(5), 136-144.
- [9] Toffler, A. (1977) *Future Shock*, Pan Books Ltd, London.



Putting information on the corporate agenda as a key resource

Sylvia E.A. Piggott

InfoPLUS, Canada

Abstract: To help us understand the trends, strategies, structure and new skills that will be required by the information professional in this global changing environment, this paper examines the future of the library and information profession and the impact that information technology — in particular the online delivery of information, reduced budget and competitive intelligence — is having and will have on this discipline. A re-engineered library model is proposed as a means of putting information on the corporate agenda as a key resource.

Keywords: project management; managerial skills; information systems; training; skills; information technology; forecasts; roles; techniques; competitive intelligence; competition; customer services; online databases; information dissemination; executives

1. Introduction

In the new knowledge driven age, where value added information is one of the most powerful and most valuable strategic tools an organisation can command, how can information professionals position themselves and structure their activities so that they can play a key role on the executive strategy team?

Today, there are dynamic forces reshaping businesses as knowledge-based, knowledge management enterprises. These forces are innovations in information technology and telecommunications which are expanding the quality, quantity and speed of communication locally, nationally and world-wide. In this environment, even corporate giants have been awakened from slumbering complacency by the intensifying forces of global competition.

In response to new and changing pressures in the competitive environment, managers must formulate and effectively implement winning competitive strategies to keep the organisation economically viable. Competitive intelligence enables managers to understand what the competition can and will do, and when and where it plans to do it. Competitive intelligence should therefore represent a key element in the strategic information management system of the organisation.

In order to achieve this competitive advantage, the organisation must have in place a formal mechanism to collect, process, analyse and disseminate competitive intelligence. The unit which naturally has the expertise to perform this job is the library or information centre. By virtue of their education, training and continuous learning, information professionals have both the skill and expertise to play a major role in scenario forecasting. The savvy information professional will know how to combine an in-depth knowledge of the information industry and its offerings with excellent management skills and professional training to command a leadership rather than a managerial role, in any organisation.

2. Skills and competencies

The information specialist who wishes to be the information leader in an organisation must demonstrate a variety of business skills in administering the library and information centre. First and foremost, the information professional must recognise that the usefulness of what we learn will be brief as things are changing so rapidly. Therefore, continuous learning will be necessary for us to maintain a leadership role.

Some of the key competencies necessary to achieve this role are:

- knowing how to communicate with computer and information technology professionals;
- knowing where hardware, software and network technology are going and positioning the library to take advantage of what is available;
- knowing how to gauge users' needs and write up a set of specifications that make technical, business and organisational sense, and do this in the context of tough budget constraints;
- entrepreneurial skills demonstrated by managing the library as a business, employing business case

methodology and showing returns on investment both in terms of quality of information and business success;

- understanding that we are moving from a paper-based, command and control hierarchy to an environment of hardware and software and documents managed by workgroups of well educated and well motivated people who no longer need much control from above;
- skills relating to organisational issues, quality and value-added services, and re-engineering;
- change management skills coupled with effective public relations management during the change process;
- negotiating skills which will result in negotiation of equitable information services contracts for the organisation.

A typical want ad for a Systems Librarian describes the ideal candidate's responsibilities as follows:

Evaluate, analyse, design, coordinate, recommend and implement equipment, software and systems for computer-based information systems and services, including Internet/intranet, CD-ROM databases, online databases, local databases, integrated library system. Maintain, develop and administer Web site, including maintenance of Internet resource collection, and develop access to library products and services via intranet. Perform system and network administration. Provide technology consulting services for colleagues, particularly Internet/intranet development, electronic resources and information retrieval system development. Develop databases of Library resources. Provide instruction for staff and colleagues in electronic resources and tools, including Internet. Provide technical support and troubleshoot local systems and software. Acts as liaison to Information Systems department.

Maintain current awareness of trends and developments in library automation and electronic resources... This ad outlines some of the key competencies which are required now and in the foreseeable future.

3. Competitive intelligence

Competitive intelligence systems will be a key resource and will be highly valued by the organisation. In the 1990s and beyond, technology will become increasingly important and organisations will tend to become more market-oriented. This will pose considerable challenges to managers responsible for the development and commercialisation of new products. Traditional approaches will not work because time-to-market will have to be reduced, product technology content will have to be increased and competitive intelligence will have increased impact on development efforts.

In today's competitive business world, informal information-gathering is no longer adequate for an organisation that wants to be proactive. So more of them are implementing systematic programmes which assimilate and interpret a broad spectrum of information on all competitors. They are known as *competitive intelligence programmes*. These programmes have nothing to do with industrial espionage and everything to do with accessing, framing and interpreting public information in more creative and insightful ways.

Given the exponential growth in information sources and the opening up of so many new global markets, the issue of business intelligence should be foremost on the strategists' minds. When enough pieces of today's business intelligence accumulate in the analytical mind, they eventually form the basis of a company's vision for the future.

This is why the world's best-run companies systematically base their decisions on a rigorous assessment of their competitive position in the marketplace and that of their competitors. Competitive intelligence, analysed and deployed with skill, can reverse an entire market share position. Business history is littered with examples of costly failures to keep a focused eye on competitors.

Competitive intelligence does not have to be costly. It is possible to do a great deal at reasonable cost simply by culling information in the public domain. In developing a competitive intelligence programme the key to success is to be focused, to know what the information is wanted for and to know what action it is intended to produce.

Published sources are a readily accessible launching pad and the information professional is adept at knowing what these published sources are and how to access them in a cost-effective way.

Technological changes, deregulation, productivity pressures, increased emphasis on quality and other factors present new problems and sources of competition. It is this highly competitive environment that appears to stimulate movement to planning styles that require externally oriented competitive intelligence. Such intelligence systems can be put in place by the information professional and will be valued by the organisation's strategic planning executives as long as these systems provide useful, actionable information to decision makers. Business intelligence is a value-added product that leads management to make a decision, undertake an action or make a deliberate choice not to act. It is not just bits of data or an unfocused competitor profile. Unless the product helps management make a decision, it does not add value.

As competition in today's crowded marketplace continues to intensify, companies find that their market research efforts need to go significantly beyond simply gathering information about customer needs and preferences. The marketplace contains not only customers but competitors, and insights about customers become useful and valuable when that information includes the appropriate competitive intelligence about what the

company's competitors are up to and what they are likely to be planning.

The only way for many companies to grow today is at the expense of their competition. If they do not, competitors will carve away their market share. A company can get a jump on the competition by knowing what the competition is up to and staying one step ahead — by employing competitive intelligence strategies. More executives are realising the need to conduct competitive intelligence operations and the information professional has a window of opportunity to provide this competitive operation for the organisation, and thereby be recognised as a key player on the strategic business information resource team.

4. New roles

4.1. Business Intelligence Strategist

A senior vice president, chief executive or product manager cares little for large mounds of data. What these people need is competitive intelligence — information that has been analysed to the point where a decision can be made. The information professional must partner with appropriate areas within the organisation to provide an organisation-wide vision for systematically collecting, filtering, analysing and interpreting relevant information necessary for valuable intelligence to be delivered to senior decision makers proactively or on an as-required basis. The new role of Business Intelligence Strategist could very well be assumed by the information professional.

A strategic business intelligence system should support the following processes:

- (1) allowing specification of impending strategic decisions, their timetable and required informational support;
- (2) determining the sources of the required information and planning for its acquisition;
- (3) obtaining all possible raw data within the available time frame;
- (4) analysing and processing the data into a client-useful format;
- (5) presenting, disseminating and storing the end product and performing a post-assignment audit.

4.2. Knowledge Manager

In the wired business environment the value of the corporation becomes the ability to generate and to effectively communicate needed knowledge throughout the system of suppliers, customers, employees and communities within which it operates. Communication, then, is everything and therefore information professionals must be experts in the effective use of the emerging information management and communication tools. This places exciting challenges and opportunities before the professional but it will also assure an exciting future.

It is estimated that as much as 80% of what a company wants to know about its competitors is known to someone somewhere inside the company, or can be found there, and so it makes sense to have a competitor intelligence structure that can be launched into action when needed. The trend in many corporations is to appoint a Corporate Knowledge Officer (CKO) who not only captures corporate knowledge but assembles knowledge on the competition as well. The new position, or at least the premise surrounding it, promises to have a tremendous impact on corporate competitiveness. The incumbent of this new position will be the savvy information professional who can demonstrate the need for such a system as well as its value. This system must be part of the organisation's risk management system. Each company needs an individualised risk management strategy because each has its own unique set of challenges.

5. Technological tools

Organisations that do not stay abreast of the latest advancements in science and technology (S&T) stand a greater chance of missing opportunities than those that maintain vigilance over the ever-changing technical environment. As a result, a resurgence of interest in technical intelligence for business is occurring in companies around the globe. Many organisations now have formal electronic intelligence programmes to gather, analyse and use S&T information to watch their competitor, to track emerging trends in technological development and to anticipate significant technology-based changes in key markets. The Electronic Intelligence System (EIS) which houses this information should be delivered to the desktops of staff in a seamless, transport manner. This kind of system is of special value to the organisation which has a significant mobile workforce since remote users can tap seamlessly into the EIS and continue to work from an informed position.

Electronic publishing and dissemination will continue to grow in importance, displacing print for most of the information used by corporate executives for business intelligence and decision-making. Therefore, much of the information services products will be delivered electronically and the library without walls will be a reality in many corporate settings. The online provider of information is challenged to supply the libraries in this setting with the necessary information which will be delivered over internal networks. Academic, public and school libraries will continue to make use of strong collections of print and other media along with physical delivery systems to serve the essential needs of their clients.

One way for the information professional to put information on the corporate agenda as a key resource is to develop a programme to build a 'knowledge management architecture' — an online competitor-and-customer database to which employees have access. This 'knowledge management architecture' system should contain continuously updated company profiles and news from public sources like Hoover, NewsEdge and other wire services; reports from salespeople and attendees' notes from conferences and conventions; an in-house directory of experts; and regulatory news, cross-indexed by company, technology and so on. All of this is information that most companies have, but it is scattered and hard to find. This system could be part of the organisation's intranet system with the information professional having the responsibility for the content of the competitive intelligence and library and information sector of the intranet.

6. The future of the information profession

The globalisation of markets is not only making things more competitive: it's making everything move faster. Companies are under increasing pressure to abandon fundamental practices such as large staff groups and slow, methodical decision-making. To survive and prosper they have implemented strategies resulting in faster, leaner and more customer-focused entities. In this kind of changing environment, information professionals must respond by being proactive so that they are not only reacting to today's realities: instead they are influencing them.

Information professionals can no longer control the speed at which information flows, or who has access to information. Clients are now accessing, via public or corporate networks, electronic services such as commercially available online services, the Internet and newswires which can communicate stories with lightning speed and visual impact.

In this new environment the role of the information professional will be changed and it will be elevated. Competition for their jobs will largely come from highly qualified technical and communications professionals. In this changing environment, the quality of the people in the profession needs to be a priority. Training programmes need to focus on elevating the standards of the profession so that information professionals can compete for and assume the leadership role in the profession. There is also a requirement to test new theories and ideas in order to redefine the information professional's and the library's roles. When this is done, the chief information professional will have an integral role in steering the organisation by managing and influencing events, not after they occur but prior to their occurring.

As the information professionals' influence increases they will be held much more accountable for the counsel they provide, the strategies they prepare and the statements they make.

To move into this era there is a need to develop best practices, in order to be recognised and accepted as professionals rather than simply as providers of a service. As part of the process, a method of evaluation must be created that enables clients and funding agencies to determine whether the work of the library has been effective.

With so many users demanding current data, information and knowledge, while demonstrating self-sufficiency in accessing information sources themselves, the information professional's department must be able to give its clients the correct tools to meet these demands while at the same time effectively managing the information resources of the organisation. When deciding on an enterprise-wide information solution, several factors should be evaluated including the size and structure of the data resources and systems, and the knowledge level of the users and the kind of information users are looking for.

The quest for value will be intensified which means that a significant amount of routine tactical activities such as the creation of information packages, or accessing research databases, will be done by the clients themselves with little or no interference from intermediaries. Clients are becoming more and more adept at doing some things for themselves as a result of improved technology.

If we are to believe the experts who peddle instant wisdom to business and governments, our future clients will be so familiar with the use of technology to access information that there will be little we can do for them or teach them.

We can't really blame the management consultants, economists, futurists or politicians for peddling this stuff because there is an inexhaustible demand for simple explanations for what is occurring and a heavy bet is being placed on high technology, the information superhighway and the wonders of the knowledge-based economy to solve our problems.

In this kind of environment we will be expected to master the technology or become marginalised. Our challenge is to differentiate between what is rhetoric and what is reality.

Information professionals have found it difficult to convince management that they represent a core function and that their efforts show return on investment. Many information professionals maintain that one can't show direct results of the work done in the information centre or library on the bottom line because there are too many subtle influences and uncontrollable variables.

It may not be easy to do so but this does not mean that dollar figures cannot be assigned to organisational problems that exist and then the changes that occur after information has been applied to the problem measured. Assuming the leadership role in the corporate competitive intelligence system is one sure way of demonstrating the value of the information professional — as well as the library — to the organisation's bottom line. It is imperative that information professionals find creative ways to demonstrate their value to the bodies which fund their operation.

7. Re-engineering the library model

Given the revolution in the telecommunications and information technologies, the globalisation of business, the growth of global networks and the improvements in storage and transmission of multimedia documents, what can the information professional do to position the library on the corporate agenda as a key resource?

I believe that to be considered as key resource a new re-engineered library model must be implemented. This model should be strategic rather than tactical, proactive rather than reactive. The model must be one where the library gets results utilising technology and managing people, and where the evaluation of its work demonstrates the usefulness of the results it produces in achieving corporate goals. The re-engineered model must embody, at the minimum, these elements:

- it must employ more sophisticated techniques for measuring the bottom line impact of its intervention;
- the value of the library must be demonstrated by becoming the hub for information acquisition and dissemination throughout the organisation. The goal of this hub would be to develop a seamless environment in which audiences that wish to can access information about the organisation's products and services; keep updated on internal and external news; monitor the competition; become informed in order to contribute to policy decisions; and engage in continuous learning. The fact that the information comes from different departments should be transparent to the clients;
- it must demonstrate outstanding business practices and fiscal responsibility. For example, rather than staffing up to achieve a goal the librarians will look for ways of forging partnerships with an appropriate area and forge a series of alliances to achieve corporate goals;
- flexibility to make the paradigm shift as opportunity presents itself. This may include shifting from a physical library to a virtual library when this is of strategic importance to the organisation;
- behave in a manner which demonstrates that the information professional moves with authority, makes decisions, takes action and takes some calculated risks which should provide strategic advantage to the organisation;
- focus on getting results and show a sense of urgency in doing so;
- exercise constant diagnosis and relentless scrutiny of internal and external environments, and be ready to take appropriate action and to make some strategic opportunities;
- perform in such a way that our clients come to believe that our professional expertise is vital to their success. Gain the client's confidence by being proactive in delivering strategic information even before the client realises that it is required;
- as funds get scarce the information professionals must give real perception that they are performing effectively and delivering good value for resources invested. It is therefore important for users to have a clear positive understanding of the value of the services performed. Publicise the advantages that clients derive from continued patronage;
- apply the principle of 'compete or perish'. Operating in this mode means applying a continuous improvement system. Continuous improvement includes continuous education;
- operate the library as though it is judged on the same basis as the clients' departments, which are given a short time to produce a given result or suffer the consequences. Keep in mind it is not whether something is being achieved but whether the clients are persuaded that it is;
- the information professional must not be content until the organisational situation is fully understood, including who the chief players are in the organisation. Whatever is important in understanding how and why the organisation produces results is of paramount importance to the information professional. Know what the organisational goals are. Identify the power centres — who are the people who get things done, who are the people who are looked up to or deferred to? Align the library with those power centres. It is politically astute this way. Learn to swim with the sharks;
- think and speak the way the key players in the organisation speak — use the keywords they use when describing what the information professional and his or her unit does for the organisation rather than using library jargon;
- provide audience segmented services. Know what each group does, thinks and feels, know what they value and what their commitments are to the organisation. How do they approach the organisation's competition? Prepare an annual report in which you talk about departments, groups or divisions which the library or information centre has assisted over the past year to achieve organisational goals;
- establish statistical benchmarks to show how the library is doing in relation to the competition. Compare numbers of staff, costs per search etc. in comparison with others in the same industry;
- when planning or budgeting for any aspect of the library's function, think of yourself as an investor for your organisation and answer questions:
 - will this activity help the organisation achieve its objectives? And how?
 - how will I demonstrate the relationship between the success of my actions and the achievement of the organisation's goals?

- what is the return on investment for a specific project? (Never talk about money spent, always talk of resources invested. Money spent is gone: resources are invested with the expectation of a return.)
- how can I document, in quantifiable terms, the status of the library's responsibilities — what benchmarks exist?

8. Benefits of the re-engineered model

There are many benefits of adopting this model of which I will highlight a few:

- you will increase the likelihood that you will be asked to serve on important committees or task forces;
- you will be more successful in the competition with other managers for your organisation's scarce resources because you are applying business methods to your requests;
- you will get authorisation for added staff when the workload requires it because you can demonstrate the need and measure the accomplishment;
- you will be far less expendable when the organisation suffers setbacks;
- the library will be recognised as a key resource for the organisation.

Working in the background, showing complacency, downplaying your professional talents, staying in your office and not mingling with key staff in different areas is one sure way of being ignored and remaining unrecognised. In addition, this kind of behaviour will ensure that the position of information leader will be assumed by a more astute business executive who will seize the opportunity to take the leadership role. The information professional must therefore be in a constant state of readiness to show leadership by assisting the parent organisation in maintaining or gaining competitive advantage.

This constant state of readiness requires a strategic approach that recognises the importance of:

- understanding how the information professional is perceived by various corporate constituents;
- creating the capacity for reaching large, diverse constituencies with precision and speed;
- finding specific solutions for specific challenges;
- looking at the short-term and long-term implications of a solution, recognising that it is not possible to predict all the long-term solutions, therefore solutions must be monitored, evaluated and changed — if necessary periodically;
- don't accept something as inevitable since we have the ability to influence the future.

9. The future in the making

Putting information on the corporate agenda as a key resource and keeping it there involves vigilance concerning what is taking place in the environment. The technology is one of the driving forces behind the collection, organisation and dissemination of information. Consequently the information professional must be on a constant watch to see what the new and proposed developments are in this area. The developments at the Media Lab at the Massachusetts Institute of Technology (MIT) are worthy of being monitored. This is especially true because corporations look to places such as these to provide leadership in predicting the future, and moreover they tend to believe these predictions!

The Media Lab at MIT, which opened about 12 years ago, promises to revolutionise life in the early 21st century in much the same way as the telephone and the automobile changed the world in the early 20th century. The guru of this lab, Nicholas Negroponte, says that doing business face to face, or even by telephone or fax, is too slow, too inflexible, too inexact — a waste of time. One of the main focuses of this Lab is to develop new ways to learn. For example, instead of dissecting frogs students will build their own computerised frogs. Instead of memorising economical or political science theories they will create their own market systems or governments. The students are therefore constructing new knowledge. In the future the answers will not be something that's waiting to be discovered: the answers will have to be invented. Teachers will become coaches in the classroom in how to use the computers to create answers.

The staff in the Media Lab are looking for new ways to blend information and artificial intelligence so that computers will do much of our thinking for us. The perfection of holography is proceeding at a rapid rate, creating free-standing holographic images. Think about information research. What many of us do now is information retrieval. We punch in keywords and the machine gives us a list of documents containing those words. But this is unsatisfactory because the retrieved list is often too long — 500 or even 15 articles will not be considered helpful to the recipient. Staff at the Lab are working on 'machine understanding' — programs that will allow computers to analyse and summarise materials for us. So, instead of 500 articles the computer will compose a one page executive summary of what has been written on the topic. The goal of the Lab is not to get people to understand computers but to get computers to understand people.

The staff in this Lab see a future where the computer will understand you and your preferences, will sort your

e-mail, package the day's news into your own personal briefings, research your entertainment options, and set-up meetings and social engagements. These computers will gather and organise information for you in the way you would yourself, if you had the time. You would in fact have a digitised personal assistant working for you in the background while you carry on with your tasks.

9.1. Narrowcasting

As the information technology progresses, narrowcasting or the possibility of preparing personalised newspapers, newsletters, news shows and magazines will be a service which will appeal to the consumer, especially the busy business executive. Individuals will be able to program their computers to extract information from a constant online stream to create information in a form and quantity that they like.

This ability to get information without the intervention of an intermediary will certainly have an effect on the livelihood of the information professional. Again, it is necessary to know what is rhetoric and what is reality in all of this. It is also important to understand the cost of implementing such systems in terms of financial outlay, as well as who will be the 'haves' and the 'have nots' in such an environment.

10. Conclusion

A well designed and well implemented competitive intelligence system coupled with outstanding library and information services is crucial to business success. If the information professional can utilise his or her competencies to implement a successful competitive intelligence system and tie it to the organisation's success, as well as implement the re-engineered library model, then he or she will be able to put information on the corporate agenda as a key resource.

Sylvia Piggott
InfoPLUS
Montreal
Quebec
Canada H4V 1Y5
Tel: +1 (514) 486 0305
Fax: +1 (514) 486 9809
E-mail: spiggott@accent.net

Bibliography

- Ashton, W.B. and G.S. Stacey (1995) Technological intelligence in business: Understanding technology threats and opportunities, *International Journal of Technology Management*, **10**(1), 79-104.
- Burke, M.E (1994) Scenario forecasting: The role of the business information professional, *Business Information Review*, **11**(1), 34-39.
- Fuld, L.M. (1996) The future of databases: What's really needed are decisionbases, *Information Today*, **13**(2), 1, 19.
- Hallaq, J.H. and K. Steinhurst (1994) Business intelligence methods — how ethical, *Journal of Business Ethics*, **13**(10), 787-794.
- Jones Jr, W.A. and N.B. Bryan Jr (1995) Business ethics and business intelligence: An empirical study of information-gathering alternatives, *International Journal of Management*, **12**(2), 204-208.
- Leake, W.W. (1996), Will the real CKO please stand up? *UE Solutions*, **28**(3), 4.
- Montgomery, G.E. and J.P. Spiers (1996) Competitive intelligence: An antidote to downsizing, *Public Utilities Fortnightly*, **134**(6), 22-25.
- Sandman, M.A and D.L. Borska (1995) Intelligence managers and intelligence representatives, *Electric Perspectives*, **20**(3), 86.
- Sorgenfrei, M. (1994) IS project management, *Systems Management*, **3X/400**, **22**(4), 10.



Disintermediation in the year 2010: using scenarios to identify key issues and relevance of IMPEL2 eLib project

Catherine Edwards, Joan Day and Graham Walton

University of Northumbria at Newcastle, UK

Abstract: 'Disintermediation' is not a term in general use, nor is it to be found in dictionaries. Among information workers it has been used to describe the diminishing role of the intermediary associated with the electronic information environment. IMPEL 2 is a JISC-funded e-Lib Project investigating the social, organisational and cultural impacts of the growing electronic library in UK Higher Education. This study is based on an earlier one (IMPEL 1) which revealed concern among library staff about disintermediation. The purpose of this paper will be to try and forecast possible developments and changes in disintermediation over the next 20 years. The structuring of scenarios has been identified as a useful tool to identify key issues. The IMPEL2 team used a scenario-building technique to identify issues surrounding disintermediation. The relevance of the IMPEL2 Project to this topic is described.

Keywords: disintermediation, scenarios, cultural change

1. Introduction

The term 'disintermediation' has crept into use only recently although the concept has been familiar to information workers since automation began. It enables the end-user to gain access to information without the assistance or intervention of an information professional. Stahl (Ref 1) points out that disintermediation is not confined to the information context but is a feature of everyday life: the use of dial telephones, the advent of self-service stores. Even books, he observes, are a technology of disintermediation, obviating the need for storytellers. Blakeman (Ref 2) refers to the ability afforded by the Internet to companies and individuals to bypass the middlemen, whether to buy insurance or an airline ticket. Birchall *et al.* (Ref 3) describe a tension between the obvious opportunities for librarianship in a post-industrial society and the professional insecurity among librarians which has increased as the era of post-industrialism has progressed. Their analysis leads them to conclude that there are two diverging directions for organisational and societal development: 'Either towards technocracy or towards the learning organisation and the learning society. In both there will be roles that could be filled by librarians, but the learning organisation's network structure would offer many more opportunities.'

Blakeman (Ref 4) is sceptical that disintermediation is a real issue. She offers a practical approach to how intermediaries may help users and how they may safeguard their future. Like Stahl (Ref 5), she recommends marketing, adding value and filtering information, using librarianship skills to maximise the quality and appropriateness of the information for the individual or organisation. The old idea of the librarian as an intermediary (some would say barrier) between user and information gives way to a new interpretation of the role.

Between 1993 and 1995 the Information Services Department (ISD) and the Department of Information and Library Management (DILM) at the University of Northumbria at Newcastle (UNN) were involved in a research project looking at the human aspects of electronic library service delivery in the UK higher education sector (Refs 6, 7 8). Eighty-two semi-structured interviews took place with a range of library and computing staff which focused on the impact of IT in their work. In many interviews with library staff, concern emerged about the perceived lack of control resulting through end-users doing their own searching (Ref 9). Other concerns from librarians about people using electronic information sources without an intermediary have been described in the literature.

In one study (Ref 5) the following conclusions were drawn from the literature about the end-user: that CD-ROM databases can be used without direction, that end-users receive satisfactory answers to their queries and that they prefer CD-ROM to printed reference tools. An analysis of 82 end-users' searches was completed to test these observations and the results complied with the assumptions. Librarians also appear worried about lack of sophistication in end-user searching. At Drexel University (Ref 11) database search print-outs/questionnaires and student bibliographies were analysed. The authors concluded that students were too specific and used insignificant terms. Another study (Ref 12) focused on whether students went beyond a basic level of searching. This was defined as entering uncontrolled search terms and displaying the result on a computer screen. Forty-six searches were analysed: only 56% used Boolean logic and only 17% used special features (e.g. command). The

conclusion was that end-users could obtain and print results but few used sophisticated refining techniques.

This lack of sophisticated end-user techniques has also been highlighted in a study at Hofstra University (Ref 13). In 1989/1990 300 questionnaires were distributed to end-users with a 68% response rate. A high satisfaction rate was reported but the librarian authors expressed their concern about the limited techniques and end-users being more interested in high recall than high precision. In another study (Ref 14) evidence was collected which showed the actual success rates of end-user searches were low but there was a high level of end-user satisfaction.

In view of the rate of technological and social change, disintermediation may become a prominent issue over the next 20 years. The purpose of this paper will be to try and forecast possible directions. Scenarios have been identified as a useful tool which can be used in a variety of way. The background will be given to the use of scenarios and the process the IMPEL2 team at UNN used in the production of the disintermediation scenarios which will be described. The key issues on disintermediation emerging from these scenarios will be examined. Finally, an overview of the IMPEL2 eLib project will be given along with its relevance to disintermediation. IMPEL2 is funded by JISC under the eLib programme.

2. Scenarios

It is important that librarians identify an evolving strategy to react to changes such as the trend towards disintermediation. Wilkinson (Ref 15) describes these kind of issues as 'long fuse, big bang'. The decisions reached about these issues will have major implications but it will take many years before it is known whether the decision was right. The 'long fuse, big bang' questions do not lend themselves to traditional analysis. It is very difficult to investigate the uncertainties on which the success of decisions will depend.

Scenarios are an attempt to deal with uncertainty. They offer 'a valuable framework for bringing together all topics being discussed and for juxtaposing them in a structure that is meaningful' (Ref 16). The IMPEL2 team recognised the value of scenarios relating to disintermediation. This paper describes the process in scenario production and the resulting forecasts.

3. IMPEL2 process in drawing up disintermediation scenario

Before the process commences a STEP (sociological, technological, economic and political) analysis of the external factors affecting disintermediation was produced. In producing scenarios the importance of the quality of information used is essential. The key role of the STEP analysis is to educate the team about the key factors which define the environment. Table 1 is the STEP analysis on disintermediation drawn up by the IMPEL2 team.

Table 1: IMPEL2 STEP analysis on external factors relating to disintermediation.

Sociological	Home working, self reliance, the learning society, changing work patterns
Technological	Growth of full text, access to individual PCs, development of CD-ROM, networking, standards, Windows
Economic	Falling prices, expense of intermediaries
Political	Superhighway, globalisation, mass higher education

There are six stages in the production of scenarios. They are briefly described below.

3.1. Stage 1: deciding on drivers for change

Here the most important factors in predicting the future environment are identified. It is important to look at least 10 years ahead. The IMPEL2 team brainstormed the disintermediation drivers for changes. It is recommended (Ref 17) that only those drivers which are 'important' and 'uncertain' are used to shape the scenarios. 'Important' and 'certain' issues should be identified but they are not to be used in the scenario-building.

3.2. Stage 2: bring drivers together into viable framework

The results of the brainstorming session were brought together to form 6-8 groupings of major drivers. The important and uncertain drives on disintermediation established by the IMPEL2 team were: government/political influence, cheap and easy access to networks/ hardware/software, commercialisation, changes in educational philosophies, polarisation of access at national and international level, breakdown of place. These were used to produce the scenarios. The team argued that the quality driver (selecting relevant information from amount available, establishing appropriateness of level) was 'important' but 'certain' and therefore would be implicit in scenarios.

3.3. Stage 3: produce initial mini-scenarios

The IMPEL2 team then brainstormed each of these drivers to produce a range of mini-scenarios relating to each driver. People were encouraged to be as inventive as possible

3.4. Stage 4: reduce to two scenarios

It is crucial that the scenarios should be complementary. The number of mini-scenarios may be reduced in number and should contain all topics. Initially five scenarios were built, then reduced to two.

3.4. Stage 5: write the scenarios

A day in the life of Professor Wheatear, year 2015

Morning

Professor Wheatear surveyed the results of her latest experiment on her hand-held PC. It was warm in the crop research environment complex today: the temperature controls were set for the crops and not for the people working there. Still she was enjoying this research and the students were enthusiastic. The work undertaken during placements like this helped them to pay for their studies at the University of the Northern Sector. She quickly added her latest calculations. No, it was no use, she simply must read that article.

She tapped in the University Library code and her SLPC number. Now the catalogue details. Yes, they had the article, but what was this? A message appeared on the screen. 'Access to this item is denied. Commercially sensitive material involved. Please apply to the Subject Librarian for clearance. Code is SUBLIBEARSCS3UNS'.

Professor Wheatear knew that access to all information sources was gained through the librarians. Subject librarians gave access to information at a price. Yet another librarian would check her credentials if the material was politically sensitive. Professor Wheatear sighed as she entered the code. This article was going to be expensive. You had to pay for everything now. It cost just to check the reference, then to read the article and even more to down-load it. 'Information is money — that's what they all said.' At last the SUBLIB was acknowledging her signal. Her Staff Level Personnel Code should be good enough to let her have access to the article, that was one advantage of being an academic. The SUBLIB needed her Personal Finance Number. If she entered it now she could read the article for ... 'Good grief — as much as that!'

Professor Wheatear decided it was worth it and entered her PFN. The screen went blank for a moment, then flashed a new message. 'Access denied. This article is legally sensitive'.

Now what could she do about that. If her suspicions were correct it was time this thing was out in the open. You would have thought the BSE scare would have taught the Ministry something but no, they still went on with the same tactics. There was no point in her trying to gain access. The system would put a stop on that and record all her entries. The Ministry were determined to keep the Greatheart findings under wraps until the court case was settled. The court case between the University of the Eastern Sector and the Ministry had been going on for some time. The article that Professor Greatheart had written on the effects of crop spraying was at the centre of the controversy and had implications for her research project. If the Ministry won then the Report might never be published and the information would be state classified. Her own work was going well but it would be so useful to check the Greatheart Project work. If she was correct, then the sooner it was known about the effects of this crop spray the better.

Phil Gates, there was the answer. A clever young man, too clever by half sometimes. That business with the hacking for instance He hadn't done any real harm, no importing viruses or dismantling systems. But he had admitted that he'd managed to get in and out of all sorts of systems undetected. He said that he believed in freedom of information and access for all. If anyone could get into the Library system it was Phil Gates. If she could persuade Phil Gates to help her ...

Afternoon

Professor Wheatear hated the drive home, in fact she hated driving altogether, come to think of it she didn't know anybody who did like driving these days. 'There's no pleasure in this anymore,' she thought to herself. Usually the Professor would avoid the rush hour traffic like the plague but after her meeting with Phil Gates she found that her concentration level was only marginally greater than that of a goldfish and so had decided to leave early. Her eyes may have been fixed upon the rear of the car in front but her mind was back at the office as she replayed the meeting with Phil Gates time and time again. What had she done? Was this article really worth risking her job for?

The traffic ahead was slowing as frustrated drivers shouted the usual torrents of abuse at each other. 'Give me strength,' she muttered to herself as she reached down into her bag on the seat next to her. 97684357. It seemed as if a full minute had passed before her daughter responded. 'Jenny, what have you been up to? Your hair's soaking!' exclaimed the Professor. 'Sorry mum, I've just got out the shower.'

'Well I finished early today dear so I'll be home in about 10 or 15 minutes.' The Professor went on 'get your hair dried, you'll catch your death!' 'OK mum, see you soon.'

The Professor pulled into the drive just as a few spots of rain began to trickle down the windscreen of the car. 'Just in time,' she thought to herself as she made a dash for the front door. She gave her daughter a big hug and asked her how her school work was going. She didn't like leaving her on her own and worked from home whenever possible but sometimes it was unavoidable.

Jenny had been working at home on a school project and Professor Wheatear was keen to hear how she was progressing. Glancing at the monitor in the front room the Professor noticed that Jenny was still online to someone called Louise in Paris and the print out on the table indicated that she had received e-mail from her contacts in Italy and Canada. Jenny's school was running a project with a number of other schools throughout the world on the subject of communication. The children from the participating schools were given each others e-mail addresses via their schools and asked to find out, using the appropriate methods of communication, as much about their hometowns as possible.

'The project looks like it's going well' the Professor said to her daughter. 'Yeah,' her daughter replied, 'and Anna suggested some articles that are online that would be useful to find out more about Vancouver, but I couldn't access them, I didn't have the correct codes.' 'Don't worry,' said the Professor 'we'll contact your school librarian and explain and see what we can sort out but first let's get some tea started, I'm starving!'

3.6. Stage 6: identify issues arising

The IMPEL2 team analysed the scenarios and highlighted the following as being important for disintermediation:

- (a) the intermediary will still exist;
- (b) the intermediary might not be a librarian;
- (c) the intermediary might have a very different role to play;
- (d) the control of the flow of information may be restricted by a number of factors.

4. IMPEL2 eLib project and disintermediation

A scenario-building exercise is fun and an excellent way of encouraging team working. Underlying the light-hearted approach in this instance, however, was the more serious intention to focus minds on important issues. The IMPEL2 Project is subtitled 'Monitoring organisational and cultural change,' and takes a campus-wide view of change associated with the increasingly electronic environment in the context of Higher Education in the 1990s. The Project is made up of four distinct but inseparable mini-projects:

- *Staff Study*: continues IMPEL1 in monitoring the impact of the electronic library on library and related support staff;
- *User Study*: runs parallel to project A in studying the impact of electronic information sources and systems on academic staff and student users;
- *Resource Based Learning Study*: considers the impact of resource based learning policies on library and information services;
- *EduLib Study*: monitors another eLib project, EduLib, which has developed a programme of educational development for library and information staff.

IMPEL2, taken as a whole, will be in an excellent position to inform the issue of disintermediation which will be central to each of the four mini-projects. It will explore the subject from the point of view of library and related (e.g. computing) staff, the point of view of library users, the teaching point of view and from the institutional and policy angle. It will focus the attention of all these stakeholders on the kind of cultural changes occurring in Higher Education, teaching and learning, and in information provision and use where technology becomes central to the activities of academic institutions. The discussion will extend beyond Higher Education to society as a whole, as technology breaks down the barriers between information seekers and sources available to them.

5. Conclusion

Most library and information professionals are well aware of the challenges facing them on a daily basis but naturally few would describe these in terms of a changing culture. To many people the concept is meaningless. However one of the aims of IMPEL2, albeit in an academic context, is to raise awareness, increase understanding and engage in discussion of the social, organisational and cultural issues which surround global electronic access to information by end-users. As the scenarios show, the issues are as much social and political as personal and individual. The use of scenarios as a means of involving groups in exploring both the wide and narrow contexts is a valuable tool. The potential for a greater appreciation of both the threats and opportunities for librarians working in a changing environment is evident.

Graham Walton
University of Northumbria at Newcastle

Online Information 96 Proceedings

Page 360

Acknowledgements

The contributions of Craig Bartle, Sally Curry, Maureen Jackson, Linda Kiakides and Margaret Watson are acknowledged.

References

- [1] Stahl, B. (1995) trends and challenges for academic libraries and information services, *Cause/Effect*, Spring, 53-55.
- [2] Blakeman, K. (1996) The future role of intermediaries, *Proceedings of ukolug96@warwick.ac.uk conference*, Warwick University, July.
- [3] Birchall, A., A. Deakin and R. Rada (1994) Knowledge automation and the need for intermediaries, *Journal of Librarianship and Information Science*, **26**(4), 181-190.
- [4] Blakeman, K. *op. cit.* (2)
- [5] Stahl, B. *op. cit.* (1).
- [6] Day, J.M., C. Edwards and G. Walton (1995) IMPEL: a research project into the impact on people of electronic libraries — Stage One — librarians, in Ahmed H. Helal and Joachim W. Weiss (Eds.) *Information Superhighway: The Role of Librarians, Information Scientists, and Intermediaries*, Essen University Library, vol. 18, 108-116.
- [7] Edwards, C., J.M. Day and G. Walton (1995) The IMPEL Project at the University of Northumbria at Newcastle: Background, Methodology and Progress, *Library and Information Research News*, **19**(2), 5-10.
- [8] Walton, G., J.M. Day and C. Edwards (1995) Training needs for staff competencies in a quality library service: relevance of the IMPEL Project, *The Liber Quarterly*, **5**(4), 389-400.
- [9] Walton, G., J.M. Day and C. Edwards (1996) Role changes for the academic librarian to support effectively the networked learner: implications of the IMPEL Project undertaken by the Department of Information and Library Management and the Information Services Department at the University of Northumbria at Newcastle, *Proceedings of the 1st International Symposium on Networked Learner Support*, 17-18 June, Sheffield.
- [10] Allen, G. (1989) Patron responses to bibliographic databases on CD-ROM, *RQ*, **29**(1), 103-110.
- [11] Wozny, L.A. (1988) College students as end-user searchers: one university's experience, *RQ*, **28**(1), 54-61.
- [12] Culbertson, M. (1992) Analysis of searches by end-users of science and engineering CD-ROM databases in an academic library, *CD-ROM Professional*, **5**(2), 76-79.
- [13] Barbuto, D.M. and E.E. Cevallos (1991) End-user searching: program review and future prospects, *RQ*, **31**(2), 224-227.
- [14] Ankeny, M.M. (1991) Evaluating end-user services: success or satisfaction, *Journal of Academic Librarianship*, **16**(6), 352-356.
- [15] Wilkinson, L. How to build scenarios, <http://www.hotwired.com/wired/scenarios/build.htm>
- [16] *Ibid.*



What to do when disintermediation looms

Cris Kinghorn

Bankers Trust Company, USA

Abstract: *An overview of the differing levels of disintermediation experienced by business information professionals with mature end-user communities, and their impact on the role of today's information centre.*

Keywords: disintermediation, reintermediation, end-user, value added, content management, knowledge management

Introduction

Being a Wednesday's child and therefore 'full of woe', when I first considered the title of my contribution to this morning's session I did initially feel myself instinctively concentrating more on the looming than on the disintermediation ... The simple truth is however that for many people disintermediation is long past the looming stage — it has been a fact of life for many service industries for quite some time. Travel agents, bankers, insurance brokers, stockbrokers, retailers, in fact most types of sales personnel across a wide variety of industries have all experienced some level of disintermediation. You can now book your holiday directly from an online kiosk in the corner of the shop, rather than watching the person behind the counter do it for you, or buy and sell stocks directly over the Internet, rather than paying your broker to do it for you. There are many examples of how the removal of the intermediary or 'middle man' is benefiting the consumer and the industry in question. Are information professionals the middle men of the information industry and therefore potentially the disintermediated? Or are they the information leaders, invaluable to end-users, guiding and controlling what they do, partnering with vendors to ensure the delivery of the right products, via the best method at the right price? The information expert that end-users will defer to as the authority with the training, the experience, the critical ability to identify the best source to provide the solution — potentially the reintermediated? It is my belief that we in the information industry are in the happy position of being able to advocate and experience many differing levels of disintermediation and turn most to our advantage.

The *Harvard Business Review* has defined disintermediation as simply 'compressing the supply chain' and it is true that greater levels have been experienced in those industries where the distribution process was perhaps more disproportionate to the whole than in others. Disintermediated industries are however also showing increasing signs of reintermediation, through the creation of new distribution channels but primarily via the redefinition of value added services.

Alternative channels of product and service distribution irrespective of the industry will force targeted marketing to niche consumers — disintermediation will cause market leaders to add value in order to retain their position as a crucial part of the process. The same *Harvard Business Review* that provided the definition (July/August 1996 — 'Nine routes to industry revolution') noted that radical developments in a particular industry or product change how they are viewed and made available. Their example was the 35mm camera — how many parents would pass over their prized Pentax to their child so that he could take photographs at a friend's birthday party? Yet one major use of disposable 35mm cameras is parents giving them to their children — the cost effective single use camera has made photography virtually universal. The traditional boundaries of that market's space were redefined by a revolutionary idea — many industries are capable of scaling down or up in order to occupy niches: survival may be dependent on them doing so.

Boundaries move

The development of end-user products, the increasing power of the desktop PC and the huge expansion in network capabilities have moved the traditional boundaries of the information world. Technology has made disintermediation inevitable for us, simply because it makes it possible — new technologies empower end-users and advance the speed of the dissemination of information. More specific data with a greater value can be given to huge numbers of people on a very cost effective basis faster than ever before. Most organisations today are putting all of their staff under pressure to work longer, harder, faster, smarter — efficient resource management

Online Information 96 Proceedings

Page 363

helps a corporation to achieve those goals. Many of the most efficient corporations are those who do not restrict access to information, whether generated internally or externally and particularly when internal and external information are integrated effectively to ensure true knowledge management. The phenomenal increase in the growth of corporate Intranets and the widespread adoption of policies such as end-user access are making the more effective use of information part of the standard business practice of more companies. Can the information professional therefore avoid disintermediation? In many ways, and particularly for those of us in the business information world, it's a little late to be considering that question. Many of us are already living with it: most of us have expanded our role to encompass the management of it.

Do it yourself

Human nature being what it is, there will always be people who will want to do what they can for themselves, either because they perceive that it's a cheaper, more timely or more flexible option or simply because they can. End-users can very easily do their own editorial trawls, find the most recent brokers' report, pull down the bio of the person they are meeting for lunch, find out how much money was spent on cinema advertising in France in the last five years, check when the patent for their competitors product expires, or the average price paid to acquire a small publishing company in Italy last year. The scenarios are endless and it is inevitable that many people will want to do some of these things for themselves: some of them may even become quite good at it. However, I think that perhaps more important than the word *inevitable* in that statement, is the phrase *some of*. Any good information centre or library is much much more than the sum of *some of*. As we are all fully aware, there is more to professional information provision than simply handing over, or e-mailing the data to whoever asked the question, or facilitating their direct access to the products and services that provide the answers. Someone has to make it their business to sift the good information products from the not so good/difficult to use/too expensive information products, to know why this service is better for this particular question than its competitors, to negotiate an acceptable price for it and to ensure that acceptable contractual conditions are in place for appropriate access.

It is perhaps in the area of content that disintermediation will fail the end-user most consistently. In a professional environment where a huge amount of editorial information is delivered electronically in a number of ways to many desks, I have regularly experienced the enquirer who insists they saw an article of precise relevance to the deal they are about to close in a particular publication and they are often able to go into ridiculous levels of detail about why they are so sure it was published here and on this date, but they have not been able to pull it from the editorial retrieval service on their desktop. Most of us will stifle wry smiles during these conversations, and undoubtedly the critical article can be found by searching the same publication but a different time period, or the same date but a different publication. How many end-users, however, will truly understand that even though they do have the correct reference they may not be able to find the article in question because although a particular file or group of files contains a publication, it may only be a relatively small proportion of that publication, or it may only provide abstracts and not the full text. The complete cover to cover full text coverage may well be in another product or simply not available electronically — good old inter-library loans or document acquisition may be the only option. End-user access may give them the ability to find most of what they need, most of the time, but without the effective content management provided by an information professional the end-user may well find themselves extremely frustrated by some of the limitations of the products on offer to them.

Doing what they do

The longer it takes an end-user to find what they are looking for, the longer they are not performing their key function — whatever it is they are actually paid to do. The bottom line is that all end-users, however good they are at finding the information they need, are actually being paid to do something else. The ability to find the data — assuming they have the will, and that is not always the case; the reluctant end-user who resents being trained to find his own exchange rate or create a table noting his portfolio's most recent market capitalisation does exist — is a skill they can only take so far precisely because it is only a part of their purpose. Even the best trained end-user may have only a patchy understanding of why that resource proves inadequate on some occasions and there is definitely a limit to the number of resources they can successfully navigate and still perform their core function effectively. Finding and delivering the most cost effective and accurate information is our core function and we should all be striving to prove the value of that function to the corporation. Sadly many information professionals still measure their value only by counting the number of requests they turn over — more requests successfully completed are perceived as better value and faster response times are perceived as better productivity. Those measurements however only tell part of the story — measuring the leverage your service provides by saving a senior manager's time so that he is free to generate more revenue also has a value and it is increasingly important in this age of disintermediation that you are able to place this kind of value on your expertise. Many senior managers in the 1995 Matarazzo survey for SLA cited as the reason for their adoption of end-user policies the fact that making the people self sufficient in the easier information tasks frees up the information centre staff to do more in-depth value added work. Twenty eight percent of the senior managers participating in that survey agreed that in-depth knowledge of information is still the exclusive domain of the information professional.

Evaluation

Hand in hand with evaluating the products and services we use and can provide to end-users comes evaluation of the end-user. He may claim to need everything you are prepared to offer him and may even suffer the training with a reasonably positive attitude, but how many will still be using all of the resources you so obligingly provide six months after desktop access becomes policy? My experience of managing direct access to a number of electronic products over several years would lead me to advise the importance of (a) training them properly and specifically where possible; (b) ensuring that they keep up to date with regular refreshers; and (c) making sure that you either run an audit programme if you are providing access via a comms server so that you can track their every move, or (d) take the time to assess each user's monthly usage and invoices so that you can establish just how effective their access is.

When we trained every new, information hungry management associate in every electronic resource, we could see usage patterns emerging within a matter of two or three months. They began to use only their favourite services, perhaps those which provided what they found to be the most attractive interfaces and intuitive navigation tools, or those which provided what they absolutely had to have in order to do their jobs properly. So, by the end of the first few months of their end-user life they almost to a man were down to using two or three online resources and two or three of the CD-ROMs. This led to some radical rethinking in our baby banker training programme the following autumn: we got more involved in the process before it began and found out which of them would be assigned an industry specialisation and which of our resources were best suited to providing exactly what they would need, applying our knowledge of our industry to our knowledge of theirs.

The regular reassessment of your end-user policy to ensure that you are providing only the most relevant services to each individual is vital. That may seem a fairly basic and obvious statement but it's advice that can often be overlooked. You may hear fellow information managers claim 'yes, we have widespread end-user access,' but how many of them ever check to see if the 25 networked at great expense users of a particular product are all still regularly using it? Getting a good price and making it work on all those desks isn't the end of the process but the beginning. Training must be both frequent and specific, allowing the end-user to attain some level of expertise in those products made available to him. The information centre with an active end-user community will find themselves playing a key role in the continuing education of that community — our role has expanded to encompass a training function yet remains the same in that we are still required to be expert in every information resource. Regularly reviewing desktop use provides valuable feedback for vendors in the development of the next generation of their product: by maintaining those lines of communication we can help in the creation of effective, easy to use and cost effective products for our users.

Varying needs

Across the globe, the information needs of a company may vary — each site needs access to resources specific to their region, to information about the markets where the divisions they work with are active, but a company marketing very similar products in a number of countries will have similar information needs at each site. End-user access in some industries is therefore increasingly becoming cross border and with the utilisation of groupware products and Intranet capabilities, several sites can now fairly easily be serviced by a single centrally located information centre. Many of the larger corporations with a number of information centres around the globe are mandating the creation of an information network to facilitate effective knowledge management. Their information centres are increasingly working together in a more cohesive fashion, usually with their IT departments, to establish standards and procedures and common information policies and provision. For example the end-user in many international banks today will have the same desktop irrespective of where he is located. One contract or several contracts with the same vendor may service these users — the information centre management is the focus of global information provision and the negotiation of global contracts for these resources. Saving your company significant amounts of money by consolidating contracts for information services and products can very positively enhance your profile and emphasise your value.

The full-service option

I mentioned a number of disintermediated industries earlier, for example stockbrokers seeing some of their clients doing their own trading. What a number of them have also reported seeing, however, is an increase in the number of clients looking for the full service option. The glut of information they can access for themselves is forcing them to become more sophisticated and to look for better long term strategies for which they need the superior knowledge of the broker. A recent survey by Arthur D. Little and the Giga Information Group attempted to measure the effect of electronic commerce on business. Almost half of their respondents did not buy the premise that disintermediation is an important reason for taking the electronic option. This led them to conclude that perhaps disintermediation is viewed as a natural result of new business practices being adopted and therefore that those who provide minimal added value will be vulnerable to disintermediation. Michael Taylor, ADL's TIME director (Telecomms, Information, Media and Electronics) was quoted as stating that only those providing 'sustainable positive value will be immune' from disintermediation and that 'reintermediation will take place as new distribution channels are formed and value added services are redefined'. A concept echoed by the

chairman of the Ontario Securities Commission, Ed Waitzer who has stated that 'even with the democratising of infrastructure and information, there will always be individuals willing to pay for someone to do it for them.'

What to do when disintermediation looms?

Well, you know it's coming, you can't avoid it and if you don't handle it in a positive way it could destroy you: lead it, follow it or get out of the way. All of us at some time or another will have to deal with some level of disintermediation. In our industry today, change is one of the few constants. I think that my job is more interesting and valuable to my employer as a result of disintermediation — we have ensured that we were part of it, that in some circumstances we advocated it and that it has become part of our remit to manage it. Part of our remit, not our sole reason for being, and not *instead of* what we did before but *in addition to*. Balancing our disintermediation with our reintermediation is one of the daily challenges which make working in today's information industry much more interesting than working in the information industry of 10 years ago.

Bibliography

Arthur D. Little, Cambridge, MA. <http://www.adlittle.com>.

Giga Information Group. <http://www.gigaweb.com>.

Harvard Business Review, July/August 1996, Nine routes to industry revolution.

Valuing the corporate library, 1995, Matarazzo, SLA. <http://www.sla@sla.org>.

Waitzer quote taken from *The Ottawa Citizen*, 16.ix.96.



Highly interactive WWW services: a new type of information sources

P. Vanouplines and P. Nieuwenhuysen

University Library, Vrije Universiteit Brussel, Belgium

Abstract: *The World Wide Web is evolving from a collection of texts linked by hypertext and hypermedia towards services that operate interactively with the information user, and which offer results through use of a broad spectrum of tools. This is demonstrated with the help of existing WWW services: a collection of interactive WWW services has been assembled and is made available through <http://www.vub.ac.be/BIBLIO/InteractiveServices.html>. The services can be classified on the basis of the client software required to access them. The evolution of the services is shown with the help of examples. In the future more services will probably produce (multimedia) results based on information located on several (distributed) servers, in a close interaction with the user.*

Keywords: Internet, World Wide Web, WWW, interactive services, active components, online access information

1. Introduction

Services on the Internet have evolved from online-accessible and CD-ROM equivalents of printed texts towards WWW hypertext and WWW hypermedia versions. Recently more reader interaction has become possible but, as a recent issue of *Nature* says (Ref 5), 'at present, Web documents are fairly static objects. You can look at them, download them and perhaps convert raw data into graphs by clicking on various plot options. But with most browsers that is about all. Indeed, the Web remains very similar to traditional publishing, with most users simply looking at information put there by others.' However, more and more that can be done on a computer will also be done on the WWW. Beyond static use lies the future of the Web: interaction.

In what follows we call the user's computer 'the client computer', or 'the client' for short; the computer that is called and on which a service resides is 'the server computer', or 'the server' for short. In fact, nothing has changed or has been added to this basic framework: at present, even the most advanced procedures for which the Internet is used still rely on the same basic principle of a user working on a client computer to obtain information from a remote server computer.

The procedures that are in now use have become more complicated on the one hand and more transparent for the general user on the other. A person who wants to get a quick taste of the WWW will have the impression that it is all very simple and easy. This impression of simplicity is wrong in a way: for instance, a question submitted to a search engine will almost always give some reply. The obtained result might even 'look good'. But a more experienced user will get 'better' results by using more adequate search engines and by formulating the request in a better way, adapted to the search engine(s). There exists a striking duality: the threshold to 'enter' the WWW has become smaller whilst becoming an expert user of the WWW is now much more difficult. We experienced this duality while we wrote this paper: subjects that are taken for granted turn out to be difficult to describe or define. But reflecting about these subjects is highly interesting. The impression of simplicity is also wrong for another reason. An inexperienced user might not be able to set up and maintain the client computer reliably: to do that, the user needs more experience. Such an effort means gaining more knowledge and spending more time. Most end-users are completely lost when something goes wrong with any technically advanced device, simply because of the high threshold. Another factor that increases this threshold concerns the variability or unreliability of URLs. All URLs mentioned in this paper were valid on 31 August 1996: no warranty can be given that they are still valid on a later date.

Let us try to define the subject of this paper in a simple way. With the help of the recent and advanced applications on the WWW we:

- sketch the evolution of the WWW;
- describe and classify advanced applications that are available on the WWW;
- discuss how the WWW might and should develop further.

2. A first look at the evolution of WWW services

After the period when only electronic mail, FTP and telnet were used, the need for structuring and organising the available information existed. Gopher, veronica and other such 'animals' came and went, to evolve towards the World Wide Web (WWW or Web). HTML, the *lingua franca* of the WWW, was created. Many frequent users of the WWW do not even know that HTML stands for HyperText Markup Language and that the magic word 'hypertext' is the core of many actions on the WWW. A user, when simply clicking on a so-called hyperlinked word or item somewhere on the client's computer screen (the contents were transferred from a server computer), invokes a new process of transfer of information from some (possibly other) server computer back to the client computer. The user knows that a marked (often called highlighted) word or term is somehow linked to other information. It is the WWW client (browser) software that reads the HTML code, that presents the interpreted code in a readable form on the screen and that translates user requests (such as the clicking on a hyperlink) to commands that are sent to the server computer.

The most recent development on the WWW concerns active components which aim at improving the ability of users to interact with WWW servers. An active component is transferred from the server to the client computer. The component executes on the client computer. The information is (often continuously) prepared and presented on the user's computer by the WWW client. The component disappears again after finishing the task. Pioneer of the active components was Sun's HotJava client, which accepts so-called Java applets. Support for Java has been integrated into other client programs: Netscape Navigator and more recently also Microsoft's Internet Explorer. The Internet Explorer also supports ActiveX active components written in Visual Basic. NCompass has developed an extension for Netscape to enable it to work with ActiveX components (http://www.ncompasslabs.com/products/Products_Main.htm). Digital proposed Oblets, which are supported by Digital's own DeckScape, WebCard and WebScape WWW clients. Oblets are distributed active components written in Obliq, which is an object-oriented scripting language for distributed computing in a heterogeneous environment (Ref 3). Lucent Technologies, a spin-off from AT&T Corporation, brought Inferno, which was developed by members of the Bell Laboratories Computing Sciences Research Center (Ref 2). Inferno is a network operating system and programming environment. Very often comparisons are drawn between Java and Inferno. Such a comparison is only partly justified because Java is a programming language while Inferno is a full operating system which contains the language component Limbo. Bell Labs provides a comparison of Inferno and Java online (Ref 1).

Table 1 indicates which WWW client software allows use of Java and ActiveX components. The table reflects the status of the available software on 31 August 1996: the situation can change rapidly. Client software which is not mentioned in the table does not accept active components. A client that is not enabled for active components may (or may not) display an indication that an active component is present in an HTML page. The information in the table is compiled from texts provided online by Microsoft (Ref 10), Netscape (Ref 12) and Sun (Ref 14).

An extensive description of Java and HotJava is made available online by Sun (Ref 9). There is also a Usenet newsgroup about Java: <news://comp.lang.java> (very much oriented towards Java developers). Interesting overviews of the HTML 3.2 specification by Microsoft (Ref 7), including specific Internet Explorer HTML tags for ActiveX components, and the official Document Type Definition by the World Wide Web Consortium (Ref 18) can also be accessed online. A more extensive introduction to the subject of interactive use of the WWW has been written by Dawes & Tregobov (Ref 7). To summarise the development of the WWW up to the advent of active components, one should distinguish between plain or straight texts and hypertexts, between the use of one medium and multimedia documents, and between a low degree of user interactivity and a high degree of interactivity (where user input determines the further development of the content of the document). This is illustrated in Figure 1 and in Table 2.

Table 1: Which WWW client software allows use of Java and ActiveX components.

Operating system	WWW client software	Java capabilities	ActiveX capabilities
Intel-based			
Windows 95	Netscape 3.0	built-in	plug-in needed
	Internet Explorer 3.0	built-in	built-in
	HotJava 1.0 preBeta1	built-in	none
Windows NT 4.0	Netscape 3.0	built-in	none
	Internet Explorer 3.0	built-in	built-in
	HotJava 1.0 preBeta1	built-in	none
Apple Macintosh			
Macintosh 68K	Netscape 3.0		none
Macintosh Power PC	Netscape 3.0		none
Unix			
DEC Alpha OSF/1	Netscape 3.0	built-in	none
HP-UX	Netscape 3.0	built-in	none
IBM AIX	Netscape 3.0	built-in	none
Silicon Graphics IRIX	Netscape 3.0	built-in	none
Sun SPARC Solaris	Netscape 3.0	built-in	none
	HotJava 1.0 preBeta1	built-in	none
Sun SPARC SunOS	Netscape 3.0	built-in	none
BSDI Unix	Netscape 3.0	built-in	none
SCO Unix	Netscape 3.0	built-in	none
Caldera Unix	Netscape 3.0	built-in	none
Sony MIPS-based Unix	Netscape 3.0	built-in	none
NEC MIPS-based Unix	Netscape 3.0	built-in	none

Table 2

Document structure	Number of media	Degree of interactivity	Example
straight	single	low	plain text file, e.g. through FTP
		high	messaging, e.g. through electronic mail
	multiple	low	viewing a video clip
		high	video conferencing
hyperlinked	single	low	HTML file, text only
		high	HTML file with forms, active components ...
	multiple	low	HTML file with links to files in other media-formats
		high	HTML file with forms, active components, ... invoking files in other media formats

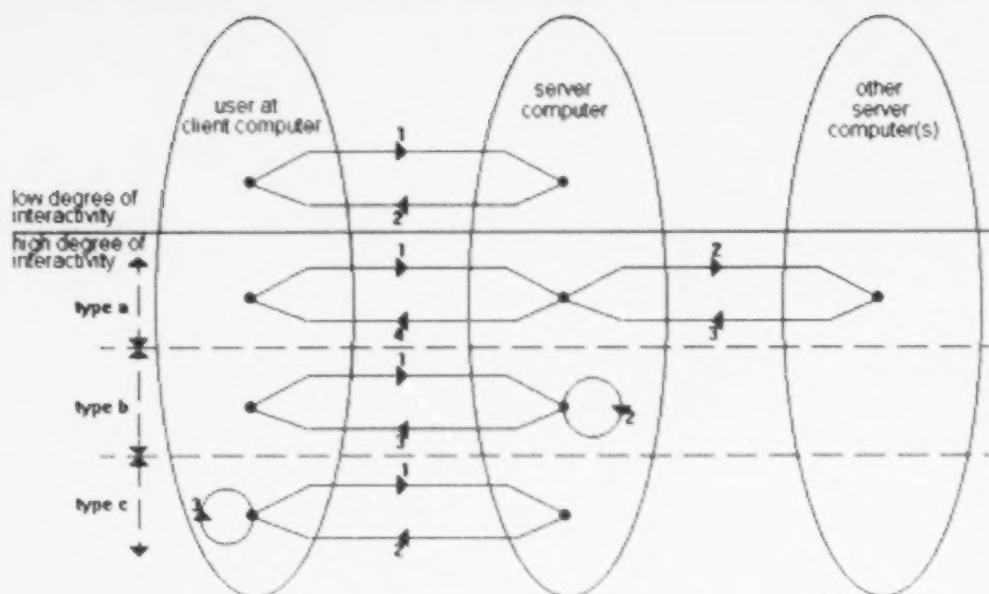


Figure 1: Types of interactive WWW services. The three types of services that have a high degree of interactivity (a, b and c) are selected. The flows of information are numbered, indicating the place in the sequence.

Udell (Ref 16) presents a view similar to ours on the evolution of the WWW up to the introduction of active components. Trupin (Ref 15) gives a how-to overview of Microsoft Internet Explorer version 3.0, Microsoft's extensions to HTML 3.2 and ActiveX.

3. Collection/selection strategy

The WWW pages discussed here are different from the majority of the material presently available on the WWW. The selected WWW services are preferably multimedia and most importantly have a high degree of interactivity, which is the subject of the paper (see Figure 2).

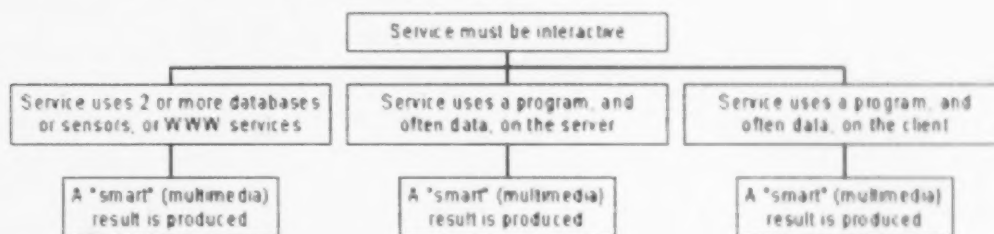


Figure 2: Scheme for the selection of a WWW service as an interactive service.

Another rather subjective way to describe the pages mentioned here is that they are 'smart' in the sense that input from the client is interpreted by the server (or active component on the client computer) and that the server sends back a response based on the client's needs or question.

We first had to look at what existed on the WWW. By a mental trial-and-error method we came to a simple set of selection criteria.

3.1. Types of low degree interactive WWW services that we do not select

- (a) Services that allow retrieval (selection) of records (elements) from a database (for example bibliographic databases, telephone directories);
- (b) Services that offer a momentary view (literally) through some instrument (camera or other sensor) (for

example a view in a room with a digital camera or a view at the contents of a soft drink machine).

3.2. Types of highly interactive WWW services that we do select

- (a) Services that address subsequent servers or tools, combine results, eliminate irrelevant or double information and thus offer a smart answer;
- (b) Services that run a program on the server computer to produce a result (for example a service that computes the result of a definite integral or a service that computes an approximation of the Mandelbrot set and sends graphical results to the client);
- (c) Services that run a program on the client computer to produce a result (the same examples of the previous point could be given here, except that the program codes are transferred from the server to the client and are executed on the client computer; this of course refers to active components).

WWW services offering combinations of the above types are possible. The three types of highly interactive WWW services are indicated in Figure 1.

Searching for these services is difficult as it is not possible to formulate a search and to submit it to Internet search engines. Alta Vista (<http://www.altavista.digital.com>) now includes a search field for the retrieval of Java applets, but not (yet) for other active components. The Yahoo catalogue (<http://www.yahoo.com>) contains fields to indicate for the presence of a Java application but these fields are not (yet) shown in the publicly accessible version (Ref 6).

For other highly interactive services one has to rely mainly on luck: other collections of interactive WWW services may contain services that satisfy our stringent criteria, one may incidentally discover a tool or somebody else may give a pointer.

4. The collection/selection of interactive WWW services

A collection of interactive WWW services that we found and selected is offered at: <http://www.vub.ac.be/BIBLIO/InteractiveServices.html>. The collection is still growing and it is clear that it will be never complete.

The following items from our collection are representative of the whole collection.

4.1. Meta search engines

A so-called meta search engine accepts a simple query formulation. The user has to specify one or several keywords: the server contacts several search engines and then offers an integrated result. This result contains a list of URLs from which duplicates have been eliminated. Two well-known WWW servers can be queried through the WWW:

- MetaCrawler Parallel Web Search Service (<http://metacrawler.cs.washington.edu:8080/>);
- SavvySearch (<http://www.cs.colostate.edu/~dreiling/smartform.html>).

4.2. HTML checkers

- WebTechs HTML Validation Service, which is more commonly known as the HALSoft validator (<http://www.webtechs.com/html-val-svc/>). This allows URLs of documents to be checked or a (small) part of an HTML document to be pasted in a form field. The document(s) can be checked against several document type definitions (DTDs) to be selected by the user;
- A Kinder, Gentler Validator (<http://ugweb.cs.ualberta.ca/~gerald/validate/>). This is a friendly, easy-to-use HTML validation service based on a real SGML parser. It is similar in function to the WebTechs validator but the returned error messages are easier to figure out.

4.3. Word and text

An overview of WWW servers about word and text was given by Pack (Ref 13). Some of the servers agree with the criteria outlined above.

- The Cyrano Server (<http://www.nando.net/toys/cyrano.html>) writes eloquent love letters and sends them by electronic mail;
- Internet Anagram Server (<http://lrdc5.lrdc.pitt.edu/awad-cgibin/anagram>): enter word(s) and it generates a list of anagrams;
- The Electric Postcard (<http://postcards.www.media.mit.edu/Postcards/>): choose a card, add your message and send it off. The recipient will be notified by electronic mail that a card can be claimed through a Pick-up Window;
- Crossword Solver (<http://www.eecg.toronto.edu/~bryn/HTML/Crosswords.html>): enter letters, with periods for unknown letters. All possible words using these letters are returned.
- Crossword Puzzle (Java applet) (<http://www.starwave.com/people/haynes/xword.html>): an interactive crossword puzzle. The puzzle is updated daily.

4.4. Science and engineering

The number of interactive WWW services related to science and engineering is by far the biggest. The following list is only a sample from the vast offering available through the WWW, which is also the case for the other sections but especially for the science and engineering section.

- EqnViewer (Java applet) (<http://www.hookup.net/~cbazza/EquViewer.html>). EqnViewer is a Java applet that allows equations to be added to a WWW page. Using a simple syntax to represent an equation one can easily insert any mathematical equation;
- Hydra-Mandelbrot/Julia fractal zoomer (<http://reality.sgi.com/employees/rck/hydra/>): zoom into the Mandelbrot or Julia fractal and graphically display the result with Hydra;
- Orientation of d-orbitals (Java applet) (<http://wwwchem.uwimona.edu.jm:1104/courses/CFTpt2.html>): Java applet that allows the orientation of the d-orbital lobes of the electrons in molecules (using wireframe representations of the orbitals that can be rotated by dragging them with the mouse) to be viewed with respect to the position of the surrounding ligands.
- Solar System Orbit Calculator (Java applet) (<http://www.humnet.ucla.edu/humnet/french/faculty/gans/java/SolarApplet.html>): dynamically shows the position of the inner (Mercury, Venus, Earth, Mars) or the outer (Jupiter, Saturn, Uranus, Neptune, Pluto) planets of the solar system. Also includes the four major moons of Jupiter, the Earth's moon, a couple of comets and a space ship that drops in from time to time (Figure 3).

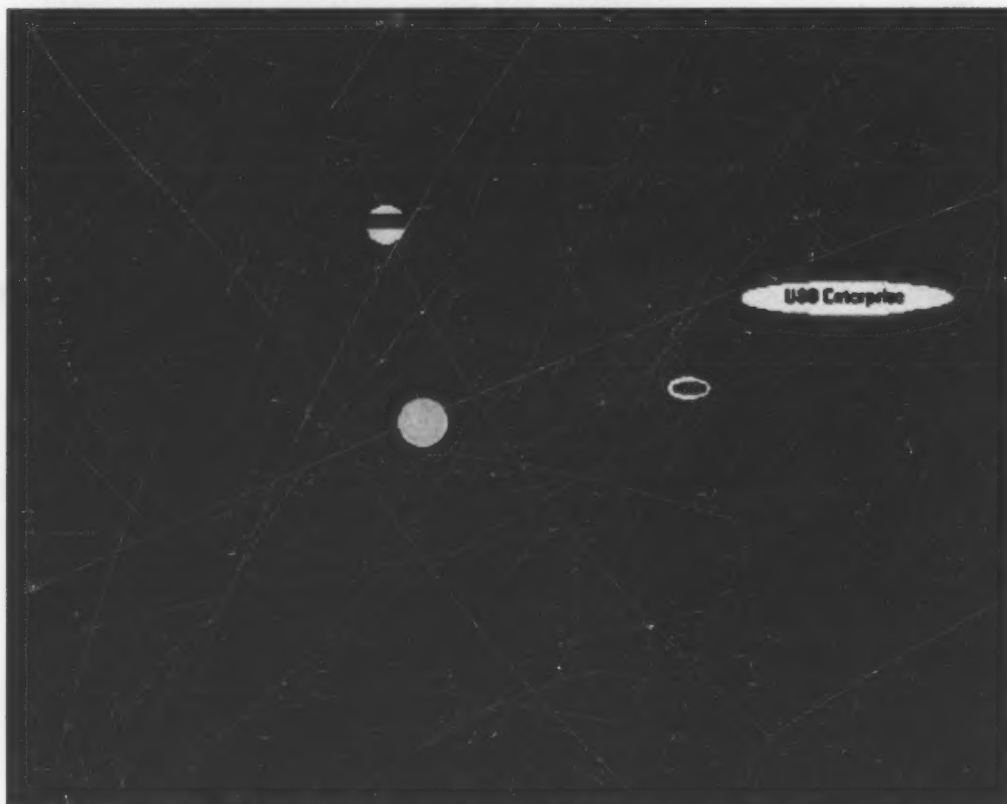


Figure 3: Screen capture of the Solar System Orbit Calculator.

- The Bradford (UK) automated remote robotic telescope (<http://www.telescope.org/rti/index.html>): the telescope and its use have already been discussed by Vanouplines & Nieuwenhuysen (Ref 17). This is our preferred example of a highly interactive WWW service. A user can set-up a complete job description for the telescope through a series of HTML forms. Help is provided in several ways by hyperlinks to classical help texts. Several tools for visualisation of the night sky are available. Once the job is submitted, the hardware and software built around the telescope monitor weather parameters, calculate the best observation time and make the astronomical observation only when all conditions are favourable. Jobs are scheduled on the basis of priorities which are mainly influenced by feedback from the user to the telescope operator (who can also raise or lower the priority level in very exceptional cases). The user is informed automatically by electronic mail upon completion of the job. The resulting picture or pictures can then be viewed through the WWW and downloaded to the user's computer where further image

processing can be performed with software that is also available from the Bradford WWW site. Resulting images are offered as GIF files or as FITS files (the preferred format for astronomical images). The Bradford telescope is in fact the only telescope available through the WWW that fulfils our selection criteria; the following two telescopes are only added to show clearly the difference — vital parts of the procedure need human interaction and in fact they offer no more than a (sophisticated) view through a device;

- The University of Iowa Remote Telescope (<http://inferno.physics.uiowa.edu/>): this telescope allows submission of observation jobs. The job description is submitted with HTML forms. The content of a job is checked manually for its content. Once approved, the observation is made. The resulting images are stored as FITS files on an FTP server. One is informed about completion with an electronic mail message;
- The University of California at Santa Barbara Telescope (<http://www.deepspace.ucsb.edu/rot.htm>): jobs are prepared and submitted with HTML forms. The jobs for this Remotely Operated Telescope are also selected and launched by human interaction. After an electronic mail notification about the job completion, the user can collect the resulting FITS images on the institute's FTP server.
- Graphical Forecast Viewer (Java applet) (<http://www-md.fsl.noaa.gov/eft/internal/gfvDemo.html>): shows some prototypes of new products which could be produced by each US National Weather Office. Interactive maps displaying weather parameters (temperature, dew point, wind speed, etc.) are created automatically using the most recent data available. One can shift to another date or select another palette set-up by clicking with the mouse in control panels.
- Spectrum Analysis (Java applet) (<http://www.tmo.hp.com/tmo/appnotes/interactive/hp-am-fm.html>): a discussion and demonstration with Java applets of the theory of AM and FM modulation including time and frequency domain representations.

4.5. Atlases and maps

- Geodetic II — World Distance Calculator (Java applet) (<http://ourworld.compuserve.com/homepages/WGiel/geod2.htm>): Java applet to calculate the distance and direction between two points on the earth. Points can be specified by entering geographical coordinates or by pointing and clicking on a world map, or a US map;
- MapQuest Interactive Atlas (<http://www.mapquest.com>): interactive street guide, with access to maps anywhere in the continental US. Allows markers to be added to indicate points of interest. There are three versions: one using forms in combination with CGI scripts, a second using a Java applet and a third using an ActiveX component;
- Online Map Creation (http://www.aquarius.geomar.de/omc/omc_intro.html): create maps interactively by filling out an input form to specify map boundaries and projection, plot locations (cities but also user defined locations), geologic plate boundaries and other tectonic features;
- Canadian National Atlas Information Service (NAIS) (<http://ellesmere.ccm.emr.ca/naismap/naismap.html>): a geographic information system. Allows multiple National Atlas spatial data layers to be viewed and manipulated to compose a map of Canada (Figure 4).

4.6. Games

- Webcube (<http://info.gte.com/fun/cube/cube.html>): a version of the Rubik's Cube. It shows two views of the cube from opposite corners to see all six faces. To turn a face one can click on it or on 18 text links;
- WebMineSweeper (<http://genesis.tiac.net/mines.html>): a version of the minesweeper game. The game starts after selecting a minefield size. Shooting is done by clicking on one of the squares;
- Puzzle (<http://genesis.tiac.net/puzzle.html>): like a toy puzzle with sliding plastic squares. A square is moved by clicking on it so that it moves to the empty space.

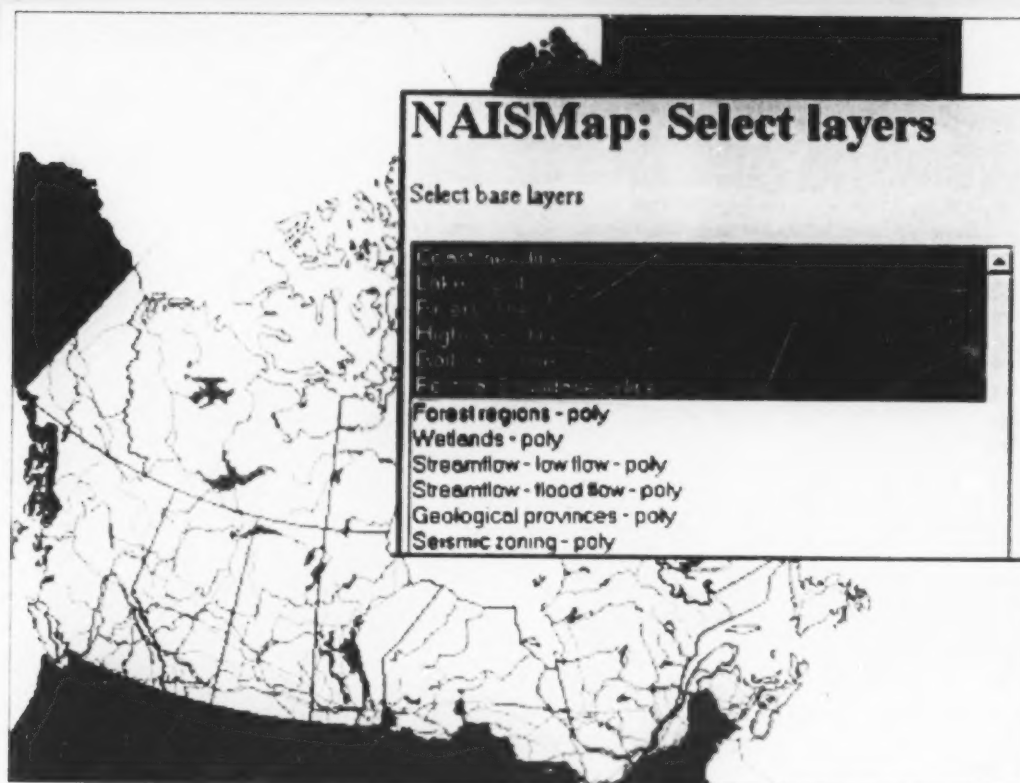


Figure 4: Example of a map obtained through the Canadian National Atlas Information Service.

4.7. Other collections of interactive WWW services

- MSN — Essentials (<http://www.msn.com/lookup/services.asp>): Microsoft offers a collection of pointers to helpful and handy WWW services. Only some of the MSN services collection pass the selection criteria presented earlier but the collection is noteworthy because of the nice form-based interfaces for each group of items. However, downloading the many forms for one group takes a long time when communication is slow;
- Gamelan — Earthweb's Java directory (<http://www.gamelan.com>): probably the most complete collection of Java components. Not that many items are highly interactive: most services transfer a simple program to the client computer, perform a simple task and display a text or graph on the client's screen. An example of a low-end Java applet is the Crossword puzzle; an example of a high-end Java applet is the Solar System Orbit Calculator (see above). The Gamelan collection also includes a list of ActiveX components;
- Applets from JavaSoft (Beta) (<http://java.sun.com/java.sun.com/applets/applets.html>): a collection of Java applets developed by Sun — a much smaller collection than Gamelan's but containing some interesting items.
- Microsoft's ActiveX Resources Area (<http://www.microsoft.com/activex/>): points users to lots of ActiveX components and to additional information on ActiveX;
- Mathematics Archives — Interactive WWW pages (<http://archives.math.utk.edu/interactive.html>): this list contains pointers to about 40 interactive WWW services that almost all fulfil the selection criteria presented here. Some of the services are small but useful tools, others are elaborate applications.
- Interactive WWW Maps and GIS (<http://www.sandia.gov/GIS/iwm.html>): a collection of WWW services with interactive maps or GIS.

5. Classification of WWW services on the basis of the client software

5.1. Services for WWW clients, following HTML version 2, without extensions

Here we find servers that are in some way the most classical (not forgetting that the WWW emerged less than ten years ago and that most of the present users did not know about the WWW a two or three years ago). Nevertheless, these servers can bring a multimedia result by including hyperlinks to text, graphics, sound and/or video files.

5.2. Services for WWW clients, HTML+ capable

User input is often through fill-in fields or forms. The information is presented to the user in text or graphical format. Sometimes a multimedia result (with graphics, sound or video) is offered. Some servers offer a continuous update of information by using the server push/client pull method; in that case one gets updates of parts of the HTML page automatically.

5.3. Services for WWW clients that accept active components

The full power of active components is not yet well exploited. One could think of 'smart' active components that partially compute results on the client computer but collect additional information from other WWW servers, to offer the final result to the user. Such a procedure could lower traffic on the Internet: (intermediate) results are calculated on the client computer and no longer need to be transferred through the Internet. This is both to the advantage of the user, who gets the results quicker, and to the advantage of other users of the Internet.

5.4. Other WWW services

Apart from the three classes sketched above, there are client/server applications that can be attached to WWW client software. It turns out to be difficult to describe and to make a distinction between the classes described above and other classes of services:

- the distinction between them is vague;
- the classes, and therefore the boundaries between them, change rapidly.

Active components and the means used by WWW clients to handle them contribute to this fuzziness. Let us coin the term *active component handlers* (ACHs) for software supporting active component. The following applications can be distinguished.

- helper application:
 - functions mostly as a standalone program to display a certain type of information;
 - called by the WWW client after the complete document has been downloaded to the client computer;
 - developed to work with one platform, with no particular WWW client software in view.
- plug-in:
 - is not a standalone program;
 - can be called while the information is still downloading (this is streaming);
 - developed to work on one platform with one particular WWW client software.
- active component handler (ACH):
 - often built into WWW client software (e.g. the Java WWW client software supports Java active components, Microsoft's Internet Explorer has built-in support for ActiveX active components);
 - but can also be provided as a plug-in (e.g. the NCompass Netscape plug-in for ActiveX components and the just-in-time Java compilers from Symantec, Borland and Asymetrix);
- active component:
 - can be handled by an ACH and is executed as an application (in this case an active component can be considered as a plug-in);
 - can be a standalone program (e.g. the extreme example of the HotJava WWW client software which is written in the Java programming language);
 - one can think of active components that, once attached to WWW client software, allow that client software to support HTML extensions that did not exist at the time that the client software was compiled and distributed.

6. Discussion

First we observe that our collection/selection of highly interactive WWW services is small, at least in comparison with the total of more than 20 million WWW documents available now. This number of services will most probably grow dramatically in a near future.

The 'information provider' should no longer be one person, or belonging to a computer centre, or belonging to an information centre. Instead there should be a close cooperation between the two so that computer-oriented persons develop technically advanced applications that contain valuable information, provided for example by skilled information scientists. The 'information intermediary' in the library still analyses the patron's request, searches commercial databases and provides the patron with an answer. The information intermediary now also includes information obtained from or through the Internet and eventually points the patron to Internet sources, where the patron himself may scan for updated information. That includes that the information intermediary may have the task to explain the use of certain elements related to the Internet, while the patron may need further help, from the information intermediary or from the computer centre (or the Internet provider), to exploit the Internet.

As we said in the Introduction, the increasing demands of the user result in more complicated systems. The systems may seem to be simpler but still in fact have a high threshold for inexperienced users. Systems may look more transparent but they are not always easy to use. For example, a person in Europe wants to download a file residing on a server in the US. The developer of the WWW page provides information about mirror sites in Europe. The person obviously selects the 'closest' server but finds out that the download time is shorter when downloading the file from the US-based server. The person in Belgium does not realise that his Internet provider in fact 'enters' the Internet at a point in the US. The user may become completely confused when pointers in a WWW page link to servers all over the world. This is only one example. Also the distinction between a client, a server and the Internet may not be clear for the user. Even if that distinction is clear, the concept of software that is transferred from the server to the client computer for execution will be new. The transparency of the WWW may turn against the user in several cases. Documents on paper are, from that point of view, easier to use and more robust.

What will the future bring? Michael Lesk (Ref 8), discussing Vannevar Bush's landmark paper in which hypertext is predicted as early as 1945 (Ref 4) touches on problems of copyright, unfulfilled wishes about a one million electronic book library, the sudden death of the WWW (like CB radio) and less freely contributed work. But Lesk states about the WWW '[that] it looks like this going to work'.

Very likely in the future there will be a combined use of server computers with active participation from the client computer. When a user sends a request to a server, the server should not have all the information resident on its storage devices. Instead it could retrieve parts of information from one or more servers, combine the results and, if necessary, get additional information from other servers, and finally offer the combined information to the user. For instance, suppose that a person wants to travel from his home town, a typical small community (thus not present in a general geographic name base), to another small town in another country. The travel must be cheap and fast, making use of public transport facilities. The WWW server should first find out the location of the two towns; next the transportation means, the travel times and the connection hours; and finally the prizes for every transportation means for each part in all of the possible routes of the travel. Then comes another task: finding the 'best' solution, i.e. quickest and cheapest. Once that is found, the possibility of ordering subway, train and flight tickets could be offered. For such a complicated task we think that it might be economical to have an active component running on the user's computer. That application should consult the subsequent WWW servers and make the combinations locally on the user's computer. By doing so, the application could show the progress of the task (which will obviously take some time), thus reducing the subjective waiting time. Also, when one or more vital servers cannot be contacted the user can be informed about this and the advice could be given to try again later, while the intermediate results are stored on the user's computer. Later on, upon re-launching the request, the stored intermediate results can be loaded again without the necessity to perform the searches on distant servers all over again.

7. Conclusion

In this contribution we have shown that an evolution takes place from classical printed text, through hypertext, towards the use of highly interactive on the WWW. Over a relatively short period, several routes have been followed by WWW developers. Thus users of the WWW have seen more movement on and heard more sound from their computers. Texts are no longer what they were for centuries: there is a marked evolution from passive reading towards active and even interactive participation of the reader, who is becoming more and more the player.

WWW clients that support active components are only emerging. Impressive novelties are lurking behind the corner. It is not so difficult to predict that active components will substantially contribute to the evolution sketched here. It is much more difficult to predict how the evolution will continue, but we may be assured that there are many more impressive developments to come.

Is there life after active components? Our answer to this question is a wholehearted 'yes', but a warning is appropriate here. New developments are coming so rapidly one after the other that one may forget the ultimate target: the user. This is the single most important target: make the WWW easier, more complete and nicer, not just more impressive. The envelope may be nice but the message in it is the most important. Agreeing with the

final sentence of Lesk's (Ref 8) paper, we say: 'If we think of information as a sea, the job of the librarian in the future will no longer be to provide the water but to navigate the ship.' We could say that many WWW developers need to change their attitude from 'It may not be useful but it is at least beautiful', to 'It is not beautiful but it is useful!'

Patrick Vanouplines
Vrije Universiteit Brussel
Pleinlaan 2 B-1050 Brussels
Belgium
Tel: +32 2 6292429
Fax: +32 2 6292693
E-mail: pvouplin@vub.ac.be

References

- [1] Bell Laboratories (1996) *A comparison of Inferno and Java*, <http://inferno.bell-labs.com/inferno/infernojava.html>.
- [2] Bell Laboratories (1996) *Inferno*, <http://inferno.bell-labs.com/inferno/>.
- [3] Brown, M.H. and M.A. Najork (1996) Distributed active objects, *Computer Networks and ISDN Systems*, 28, 1037-1052.
- [4] Bush, V. (1945) As we may think, *Atlantic Monthly*, 176(1), 101-108. Also available through <http://www.isg.sfu.ca/~duchier/misc/vbush/vbush-all.shtml>.
- [5] Butler, D. (1996) Research and the Internet connection, *Nature*, 380, 377-381.
- [6] Callery, A. (1996). Yahoo! Cataloging the Web, *Proceedings of the Conference 'Untangling the Web'*, 26 April 1996, University of California, Santa Barbara, <http://www.library.ucsb.edu/untangle/callery.html>.
- [7] Dawes, H. and A. Tregobov (1996) *The Interactive WWW and its Use in Education*, BSc Computer Engineering Thesis, University of Manitoba.
- [8] Lesk, M. (1995) *The Seven Ages of Information Retrieval*, <http://community.bellcore.com/lesk/ages/ages.html>.
- [9] Goslin, J. and H. McGilton (1996) *The Java(tm) Language Environment: A White Paper*, http://java.sun.com:80/doc/language_environment/.
- [10] Microsoft (1996) *Internet Explorer 3.0 Comparison Guide*, <http://www.microsoft.com/ie/ie3/compare.htm>.
- [11] Microsoft (1996) *HTML Reference*, <http://www.microsoft.com/workshop/author/newhtml/default.htm>.
- [12] Netscape (1996) *Netscape Navigator 3.0 Reviewer's Guide. Version 1.0*, http://home.netscape.com/comprod/products/navigator/version_3.0/review.html.
- [13] Pack, T. (1996) Electronic words — a word lover's guide to digital dictionaries, thesauri, and other cyberplaces, *Database*, 19(2), 25-31.
- [14] Sun (1996) *HotJava FAQ*, <http://java.sun.com/HotJava/hjfaq.html>.
- [15] Trupin, J. (1996) Microsoft Internet Explorer 3.0 — an authoring overview, *Microsoft Interactive Developer*, 1(3), 18-39.
- [16] Udell, J. (1996) Your business needs the Web, *Byte*, 21(8), 68-80.
- [17] Vanouplines, P. and P. Nieuwenhuysen (1995) Online information about astronomy: the Chiron case, *19th International Online Information Meeting Proceedings*, London, 5-7 December 1995, Learned Information Europe Ltd, Oxford, pp. 311-325.
- [18] World Wide Web Consortium (1996) *Introducing HTML 3.2*, <http://www.w3.org/pub/www/MarkUp/Wilbur/>.



OMNI — alternative approaches to Internet metadata

Sue Welsh

The OMNI Project, National Institute for Medical Research, UK

Abstract: *The growth in size of the Internet has resulted in much effort being spent on indexing its contents. The most popular solutions are created by automatic methods and although offering impressive coverage they disappoint where precision of meaning is required. Alternative services created by human beings arrange and index resources according to concept and offer more relevant retrieval, but cannot hope to achieve 100% coverage. New developments are imminent which may support both the automated and non-automated approaches, and vastly improve the quality of Internet metadata.*

1. Introduction

The quality of information on the Internet is today as much a talking point as the Internet itself. The World Wide Web (WWW) in particular has given everyone a chance to publish, and finding what you need in the resulting free-for-all is becoming increasingly difficult.

Why should this be when there is no shortage of tools designed for this task? The most comprehensive solutions are, by necessity, computer generated. This total approach to Internet indexing as practised by Alta Vista (Ref 1) and Lycos (Ref 2) is in sharp contrast to the subject-based information gateways (SBIGs) such as OMNI (Ref 3), SOSIG (Ref 4) and EEVL (Ref 5), and electronic communities such as PharmWeb (Ref 6) and BioMedNet (Ref 7). In between these two extremes, indexes and directories abound. Why aren't we being well served with all this choice?

In an article entitled 'Seek and Ye Shall Find (Maybe)' in the Internet style magazine *Wired*, Steve Steinburg joked

'... information retrieval is really only a problem for people in library science. If some computer scientists were to put their heads together, they'd probably have it solved before lunchtime' (Ref 8).

but the problem is clearly still with us despite occupying some of the Internet's brightest minds for several years.

This paper will seek to demonstrate that the task of locating information on the Internet has not been and will not be accomplished by the use of sheer computing power alone. It will review the topic of metadata creation with special reference to the OMNI project and look forward to two encouraging new initiatives: PICS and the Dublin Core/Warwick Framework.

2. Metadata, what is it and why is it necessary?

2.1. What is metadata?

Metadata is data about data. Its most familiar form is the descriptive catalogue record used to describe printed books, as defined by AACR2 (the Anglo American Cataloguing Rules) and expressed in MARC format.

2.2. Metadata formats

In contrast to the very well defined standards in existence for printed material, there is no standard form of metadata for Internet objects. Projects creating it have used various formats and some have created their own: for example NISS (Networked Information Systems and Services) (Ref 9) developed the NISS template for its gateway to networked resources (Ref 10).

OCLC has used MARC-like records in NetFirst (Ref 11) which has recently been made available to the academic community in the United Kingdom by CHEST.

Metadata can range far wider than the traditional descriptive catalogue record; administrative data (e.g. pricing, administrative contacts, authentication instructions) and content ratings (e.g. stars or marks out of ten) are two such examples.

2.3. Automated metadata

Both NISS and OCLC create rich metadata records by relying on human input. Automated indexes also create metadata but are far more limited in this respect as:

- they rely on the internal structure of the HTML (HyperText Markup Language) document;
- they cannot glean any information from non-textual objects such as images or sound;
- they are hampered by poor practice from HTML authors, e.g. lack of <title></title> in the head of the document;
- they can be fooled by simple tricks such as word spamming (the practice of including multiple repetitions of a keyword in a HTML document in such a way that is invisible to the person viewing with a standard Web browser but picked up by crawlers/robots. It has been used by unscrupulous Web authors to ensure their documents float to the top of a ranked list of resources).

Even putting these problems aside, retrieval from automated indexes suffers because they index the full text and not concepts (but see Excite later.) This means that false drops due to homonyms (a single word with multiple meanings) are common and reasonably comprehensive searching is likely to require knowledge of too many synonyms (multiple words with the same meaning) to be possible. This makes automated search engines good for searching for a very specific needle in a large haystack, but poor for searches on topics with any breadth.

2.4. Automated concept indexing

Can concept indexing be automated? The makers of Excite (Ref 12), one of the more innovative search engines would like us to think so. 'Excite ... has the intelligence to search through reams of information and bring you the good stuff,' claims the publicity, but in fact there is no intelligence at work here, just more computing power applied with flair. Using its full text index, Excite clusters documents that have similar profiles (many keywords in common) and creates a classification scheme as a result. How this affects the quality of results returned probably depends on the type of search but false drops and inexplicable hits seem to occur with Excite as with the other automated indexes. At the present time, content indexing is firmly in the hands of the AI researcher.

In a nutshell, automated indexes fail to satisfy all our search requirements because they use processing might to take the place of a decent index or high quality metadata.

3. Creating metadata, a case study

How can we create a better index? If robots are not good at creating metadata, can it be created for the Web by hand? A group of projects under the eLib (Ref 13) umbrella are investigating a different model. A case study of the OMNI project follows, illustrating the processes and problems involved.

3.1. Background to the OMNI project

The OMNI project is part of the Access to Networked Resources group (ANR) of the Electronic Libraries Programme (eLib). OMNI is compiling a gateway to high quality biomedical information available via the Internet and also produces a comprehensive listing of UK sites in this area. Funded by the UK's Higher Education Funding Councils for a two year period, OMNI began work in July 1997 and launched the first edition of its service in November 1995.

3.2. The OMNI philosophy on metadata

We have seen that the most widely used search and retrieval tools are unsuited to certain types of search query because of the way in which their metadata is constructed. This is particularly true of subject-based inquiries. They also offer no quality control over resources included: indeed it is their aim to include everything. The challenge faced by OMNI was to seek an alternative approach, which must differ from what existed previously by:

- (1) being tailored to the requirements of the research and academic community;
- (2) offering access to resources in one subject area, biomedicine;
- (3) selecting on the basis of quality;
- (4) providing metadata that would effectively guide users to the right resource, eliminating the need for directionless 'surfing';
- (5) paying special attention to the needs of the UK community.

Each of these five key issues was also a point against the use of the traditional automated methods. Neither restricted audience focus, subject requirements or the effect of geographical location on user needs can be expressed in such a way that makes it possible to program a robot. While it is possible to direct an automated process towards resources in the academic/research domains (e.g. .ac.uk or .edu) or the UK domain alone (.co.uk, .ac.uk, .gov.uk etc.) this is an oversimplification of the requirements expressed above. Serving the information needs of the UK community is not the same as giving access to all that that community publishes on the Web, neither does collecting together the combined outputs of the various educational domains provide all that

that community may require.

Quality is also in the eye of the beholder. There is no way to assess it other than establishing criteria which must be judged by a real person: a robot cannot judge content, usability, intended audience and so forth.

Finally the kind of metadata that is required was not available, and is not yet present in large quantities, to be picked up from the Internet itself. Robots can only pick up what is there. Compare and contrast the OMNI metadata for the home page of the Center for Complex Systems Research, at the Beckman Institute for Advanced Science and Technology (University of Illinois at Urbana-Champaign), with the automatically created entry from a major search engine (Figure 1). A rather extreme example which serves to illustrate many of the problems with automatically gathered metadata!

OMNI metadata (selection)

Title: Center for Complex Systems Research, University of Illinois at Urbana-Champaign.

Description: The home page of the Center for Complex Systems Research (CCSR), at the Beckman Institute for Advanced Science and Technology (University of Illinois at Urbana-Champaign). The CCSR is an interdisciplinary group investigating a variety of complex dynamic processes including biological intelligence (endodynamics, neural networks and metabolic networks). This site offers access to background information about the Center and the full text of technical reports generated by the group.

Keywords: neural networks (computer); acclimatization; metabolism

Accessible by: <http://www.ccsr.uiuc.edu/>

Automated metadata

No title

This page has moved to here.

<http://www.ccsr.uiuc.edu/CCSRHome.html> - size 132 bytes - 27 Mar 96

Figure 1: Internet metadata.

Having decided that human intervention was essential it remained to choose a metadata format to work with. There is no standard for Internet metadata, although there are several potential candidates which may become the standard of the future. One of these is the IAFA (Internet Anonymous File Archive) template, which was originally proposed as a method of describing the contents of FTP servers and now includes formats for a variety of other Internet objects.

3.3. ROADS and IAFA

In common with the majority of ANR projects including the established and well regarded SOSIG (social science) service, OMNI chose to use the ROADS (Ref 14) database software. ROADS (Resource Organisation and Discovery in Subject-Based Services) is another eLib project developing a user-oriented resource discovery system. ROADS is based around the IAFA (Internet Anonymous FTP Archive) metadata format (Ref 15), illustrated in Figure 2. IAFA records consist of a list of attribute:value pairs and exist for various types of Internet objects, software, services, documents, images etc. Although ROADS is committed to developing IAFA it recognises that other formats may become more important during the lifetime of the project and interoperability must be ensured.

OMNI is currently using the first version of the ROADS software which allows creation, editing, searching and browsing of IAFA records. Future ROADS developments will facilitate simultaneous access to distributed ROADS databases using the WHOIS++ directory service protocol.

Template-Type: DOCUMENT

Handle: 814014233-2699

Category:

Title: Medical/Clinical/Occupational Toxicology Professional Groups

URI-v1: <http://www.pitt.edu/~martint/pages/motoxorg.htm>

URI-v2: <http://www.pitt.edu/~martint/pages/motoxorg.htm>

Admin-Name-v1: Thomas G. Martin

Admin-Work-Postal-v1:

Admin-Job-Title-v1:

Admin-Department-v1:

Admin-Email-v1: TGM@MED.PITT.EDU

Admin-Handle-v1:

Admin-v1:

Description: A list of (mostly US) professional groups of interest to toxicologists. A short description of each organization is given, as well as addresses, phone numbers and contact names.

Citation:

Publisher-Handle-v1:

Publisher-Name-v1: Toxicology Treatment Program

Publisher-Postal-v1: University of Pittsburgh Medical Center,
Pittsburgh, Pennsylvania, USA

Publisher-City-v1:

Publisher-State-v1:

Publisher-Country-v1:

Publisher-Email-v1:

Publisher-Phone-v1:

Publisher-Fax-v1:

Keywords: toxicology; poison control centers

Version-v1:

Version-v2:

Format-v1:

Format-v2:

Size-v1:

Size-v2:

Language-v1:

Language-v2:

Subject-Descriptor-v1: 615.9

Subject-Descriptor-v2: QV600

Subject-Descriptor-Scheme-v1: UDC

Subject-Descriptor-Scheme-v2: NLM

Destination: OMNIworld

Record-Last-Modified-Date: Thu, 18 Oct 1995 11:06:31 +0000

Record-Last-Modified-Email: unknown@nimp377.nimr.mrc.ac.uk

Record-Created-Date: Thu, 18 Oct 1995 11:03:53 +0000

Record-Created-Email: unknown@nimp377.nimr.mrc.ac.uk

Figure 2: An IAFA record from the OMNI database.

3.4. Creating quality metadata

All OMNI metadata is created by hand and occupies three members of staff and a band of over 80 volunteers from the bioscience community. By the end of its initial period of funding the project aims to have 2000 completed records. Not all the IAFA attributes are used and only four are displayed to the public (title, description, keywords and URL) at present.

A record entering the OMNI database follows the following path:

- discovery e.g. announcement via e-mail or newsgroup, serendipitous discovery or as part of organised scanning of major sites;
- evaluation according to OMNI's guidelines, which are available for public scrutiny and comment;
- description: either a partial record is composed by an OMNI volunteer and completed and checked by OMNI staff or the entire record is authored by OMNI. Descriptions include a summary of the content of the object, MESH (Medical Subject Headings) indexing and classification by NLM (National Library of Medicine) and UDC (Universal Decimal Classification);
- dissemination: OMNI users are alerted periodically that new resources are available by e-mail and via the WWW.

This is clearly a time consuming process and OMNI is a service at the opposite end of the scale to Alta Vista and Lycos. The subject-based information gateways in general are characterised by:

- small, well defined areas of subject interest;
- selection of resources based on quality;
- added value in the form of description, indexing and classification.

Only time will tell if this approach can fill the gaps left by the automated search engines. Meanwhile, some exciting developments might make the task of providing both types of service easier.

4. Future developments

Human intervention results in better metadata but is time consuming. The amount of information flooding into the public domain via the Internet is so great that services like OMNI can never hope to create rich metadata records for every document within their subject area: they must select using some set of criteria (in OMNI's case selection is made following assessment of the quality of the object according to our own quality criteria.) Selection itself, though, takes more time. How can this whole process be speeded and what can we do about the resources which will never be considered important enough to catalogued in this way?

4.1. The Dublin Core

A key initiative attempting to answer these questions has put forward suggestions for author-generated metadata. The Dublin Core Metadata Element Set, proposed at a meeting in Dublin, Ohio in 1995 (Ref 16) and consolidated at a second conference in Warwick, UK (Ref 17) is a set of 13 key metadata elements for describing document-like objects on the Internet (see Figure 3.)

Subject: The topic addressed by the object

Title: The name of the object

Author: The person(s) primarily responsible for the intellectual content of the object

Publisher: The agent or agency responsible for making the object available

Other Agent: The person(s) such as editors and transcribers, who have made other significant intellectual contribution to the work

Date: The date of publication

Object Type: The genre of the object, such as novel, poem or dictionary

Form: The data representation of the object, such as Postscript file

Identifier: String or number used to uniquely identify the object

Relation: Relationship to other objects

Source: Objects, either print or electronic, from which this object is derived.

Language: the Language of the intellectual content

Coverage: The spatial locations and temporal duration characteristic of the object.

Figure 3: The Dublin Core, taken from <http://cs-tr.cs.cornell.edu/~lagoze/warwick/warwick.doc>.

The Ohio meeting, which first proposed the 13 elements, was organised by OCLC and NCSA (the National Center for Supercomputing Applications). The Warwick Conference (sponsored by UKOLN, the UK Office for Library Networking) took this simple idea forward by suggesting a mechanism for carriage of metadata (which may be Dublin Core or some other metadata set). Of most immediate interest is a specified special case where metadata is embedded in the HTML document it describes. Put simply, this offers Web authors the chance to catalogue their own material in a standard format. It will then be a simple matter for a robot to pick up this metadata and pass it onto an index. Automated indexes would be able to gather in better metadata and authors achieve better visibility for their work; selective sites can automate some proportion of their metadata creation; everyone benefits.

The Dublin Core will have a major impact on Web indexing only if authors and publishers can be persuaded to use it, and there are proposals to incorporate a mechanism for easy addition of the required data into common HTML authoring tools. The accuracy of author generated metadata is another problem altogether and it is likely we will continue to see services like the subject-based information gateways employing editors to check the veracity of the data being harvested.

4.2. The PICS standard

The PICS (Platform for Internet Content Selection) (Ref 18) standard is designed to facilitate provision of a specific type of metadata — content rating. The idea of scoring or ranking Internet sites by some agreed scheme is not new. So called evaluative search services such as Magellan (Ref 19) are already doing just that. Magellan reviewers award sites a marks out of ten for depth, ease of exploration and 'net appeal', combining these scores to produce an overall star rating of one to five stars. When you browse or search using Magellan your results will be returned with star ratings.

Evaluative services seem at first to be an excellent way of finding good material but are aimed at a popular, general audience (the concept of 'net appeal' is not useful to the average clinician, for example). Their major problem lies with the breadth of the material on the Web and the variety of people using it. How well does a single set of rating criteria deal with the range of information available? How can a single reviewer award a rating that makes sense to everyone from Star Trek enthusiasts to molecular biologists? Surely we need different ratings for different subject areas and communities? Again we return to a key characteristic of this information we call metadata. Each community has a clear and different idea of what metadata is and each community is right: your idea of useful metadata depends on the data you use and what use you make of it.

The avowed purpose of the PICS standard is to stem the rising tide of concern amongst parents that their children are accessing unsuitable material via the WWW. The idea is simple: PICS compliant browsers will point at Web resources via a ratings label. The label may be published by a trusted third party providing ratings information for a particular grouping in the form of a database, or information providers may label their own resources. Parents will be able to set up their browser to deny access to material without a suitable rating. Parents with different attitudes to 'offensive' material will be able to point at different trusted third parties or deny access to different levels of material. PICS has been described by its proponents as value-neutral, as it enables the implementation of context-specific access rather than global restrictions which they believe would destroy the freedom of expression that makes the Internet so popular.

PICS technology is being developed by the World Wide Web Consortium and (perhaps even more significantly) has won the support of industry leaders such as Netscape, Microsoft, AOL (America Online) and CompuServe. At the Fifth International WWW Conference in Paris this year a PICS compliant version of Microsoft's Internet Explorer was demonstrated and other browsers will not be far behind. It is less clear who will provide labelling services, but there are clear commercial opportunities and some sites are already offering PICS compatible rating systems.

5. Conclusion

Publishers on the WWW have a vested interest in making sure their site is visited. This may be because they have information there that they need people to see, or because they earn money through advertising and sponsorship (dependent on access statistics), or because they wish to sell you something themselves when you arrive.

Librarians and other information workers face a future where they must coordinate the old and new information domains and guide their users, as they have in the past, to information which is timely, accurate and conveniently located and priced.

Users, if they don't already, will want to use the Internet to save time rather than (as is now only too often the case) waste it.

All of us need improved search tools with which to navigate the Internet. I believe that the answers to our present dilemma will be based on more than one tool, on both human intervention and intelligent automated data gathering. Metadata, as exemplified by the fusty old catalogue card, is set to become a key concept in the future of the new information world.

Sue Welsh
OMNI Project Officer
National Institute for Medical Research
The Ridgeway
Mill Hill
London NW7 1AA
UK
E-mail: swelsh@nimr.mrc.ac.uk

References

- [1] *Alta Vista Search* [online]. Digital Equipment Corporation. Available from <http://www.altavista.digital.com/> [accessed 1 September 1996].
- [2] *Lycos Search* [online]. Lycos, Inc. Available from <http://www.lycos.com/> [accessed 1 September 1996].
- [3] *OMNI: Organising Medical Networked Information* [online]. OMNI Consortium. Available from <http://omni.ac.uk/> [accessed 1 September 1996].
- [4] *Social Science Information Gateway — SOS/IG* [online]. Bristol University: Centre for Computing in the Social Sciences. Available from <http://sosig.ac.uk/> [accessed 1 September 1996].
- [5] *Edinburgh Engineering Virtual Library (EEVL)* [online]. EEVL Consortium. Available from <http://eevl.ac.uk/> [accessed 1 September 1996].
- [6] *PharmWeb* [online]. Manchester University. Available from <http://www.mcc.ac.uk/pharmweb/> [accessed 1 September 1996].
- [7] *BioMedNet* [online]. London: Electronic Press. Available from <http://www.biomednet.com/> [accessed 1 September 1996].
- [8] Steinburg, S. (1996) Seek and ye shall find (maybe), *Wired*, May.
- [9] *NISS Information Gateway* [online]. University of Bath: NISS (National Information Services and Systems). Available from <http://www.niss.ac.uk/> [accessed 1 September 1996].
- [10] *NISS Directory of Networked Resources* [online]. University of Bath: NISS (National Information Services and Systems). Available from <http://www.niss.ac.uk/subject/index.html> [accessed 1 September 1996].
- [11] *NetFirst* [online]. OCLC Online Computer Library Center, Inc. Available from <http://www.netfirst.ac.uk/> [accessed 1 September 1996].
- [12] *Excite* [online]. Available from <http://www.excite.com/> [accessed 25 July 1996].
- [13] *eLib: The Electronic Libraries Programme* [online]. University of Bath: eLib. Available from <http://ukoln.bath.ac.uk/elib/intro.html> [accessed 1 September 1996].
- [14] *ROADS, Resource Organisation And Discovery in Subject-Based Services* [online]. Loughborough University of Technology: Department of Computer Studies. Available from <http://www.roads.lut.ac.uk/> [accessed 1 September 1996].
- [15] Deutsch, P. *et al.* (1995) Publishing Information on the Internet with Anonymous FTP [online]. Available from <http://info.webcrawler.com/mak/projects/iafa/iafa.txt> [accessed 1 September 1996].
- [16] Weibel, S. (1995) Metadata: the foundations of resource description, *D-Lib Magazine* [online], 7. Available from <http://ukoln.bath.ac.uk/dlib/dlib/July95/07weibel.html> [accessed 1 September 1996].
- [17] Lagoze, C. *et al.* (1996) The Warwick framework: a container architecture for aggregating sets of metadata. Available from <http://cs-tr.cs.cornell.edu/~lagoze/warwick/warwick.doc> [accessed 1 September 1996].
- [18] *Platform for Internet Content* [online]. World Wide Web Consortium. Available from <http://www.w3.org/PICS/> [accessed 1 September 1996].
- [19] *Magellan Search* [online]. McKinley. Available from <http://www.mckinley.com/> [accessed 1 September 1996].



Smart systems, smart searches

Bette S. Brunelle

Ovid Technologies, USA

Abstract: Almost overnight, the World Wide Web has made the solution of 'classic' information retrieval problems a pressing commercial goal. Interestingly, in spite of a prevalent attitude summarised in a recent issue of *Wired* that 'information retrieval is really only a problem for people in library science — if some computer scientists were to put their heads together, they'd probably have it solved before lunch time,' the various I/R solutions being offered on the Web are quite familiar to the librarian. The various Web search sites make use of 'traditional' inverted indexes, manual indexing, automatic indexing based on statistical models, relevance ranking and document clustering. All of these statistical techniques have been used, usually along with a 'traditional' Boolean Search engine, in various commercial information retrieval software products, and each has its strengths and drawbacks. This paper will examine some of the strengths and weaknesses in terms of Web searching, and close with some observations on the current state of affairs and predictions for the future.

Keywords: Information retrieval, statistical techniques, search techniques, advanced searching, relevance ranking, thesaurus, document clustering, Internet, indexes

1. Introduction

Some of the most-often used sites on the Web are those such as Yahoo!, InfoSeek, Excite and Alta Vista that provide search engines and indexing systems which ease the process of navigation on the Web. Interestingly, in spite of the optimistic attitude that 'information retrieval is really only a problem for people in library science — if some computer scientists were to put their heads together, they'd probably have it solved by lunch time,' (Ref 1) the various I/R retrieval solutions on the Web are simply implementations of various Boolean and statistical techniques that have already been developed by computer scientists, and much discussed at librarian conferences, for the past twenty years. As with any I/R indexing schemes or retrieval systems, each has its strengths and drawbacks — and none can claim to have 'solved' I/R.

A survey of a few of the Web search systems corresponds to a summary of the major search indexes and statistical techniques currently available. As will become clear, all of them are missing features which are common with CD-ROM or 'traditional' online systems. The systems do, however, break ground in attempting to let statistical processes and 'behind-the-scenes' manipulation of user input substitute for some of the set manipulation with which professional searches are accustomed to refining searches. Because the Web is a stateless system lacking the notion of a session, building on prior search sets is not possible. Thus the Web — by its nature a casual, end-user system — is the ideal testing ground for search techniques that by-pass complex Boolean logic.

2. Searching on the Web

A number of the Web search services rely on the most-familiar professional I/R retrieval system of all — the inverted index. 'Traditional' Boolean systems are based on the inverted index — all significant (non-stopwords) words in a document (in this case a Web page or part of a Web page) are posted in a list (the index) which includes the word and a pointer to the document which contains it. Some indexes also show which words are adjacent or near to each other in a document.

Unlike traditional inverted-index databases, which must store not only the index but also the actual text for each document, Web search indexes only have to store the index and a pointer to the document — the document is stored out on the Web somewhere and can easily be accessed separately from the search system. Because of this the storage requirements for Web databases are not as great as for the traditional ones but, given the size of the Web, the inverted indexes themselves can quickly become very large scale — at the current 20% growth rate, by 1998 a complete Web inverted index will exceed the almost 29 terabytes of information in the current US Library of Congress (Ref 1). What performance limitations there might be to indexes of this size is unknown.

Of course, at present none of the Web search systems claim to index the entire Web and it is difficult to know exactly what is indexed from service to service. Some index only Web pages; some include Listservs and Usenets. Some claim to index every word on a page and some index only the titles, keywords and/or the first paragraph. There is often no notion of fields corresponding to the document structure. Even where some field

structure is offered, the fields have to do with the structure of the Web (title, URL, host) rather than with the structure of a particular document. Distinctive data elements of a type of document, say 'book reviewer', are not available, nor are the kinds of delimiters often made from these data elements, for example 'book review.'

The Boolean sites typically offer AND/OR/NOT searching, although some have more extensive operators. Rules of entry vary from system to system and are often found only in help text. Many Boolean sites use techniques such as automatic truncation to increase retrieval — this can be both bewildering (when retrieval does not match input) and frustrating (when an already high noise level is increased).

Other Web search systems do not allow Boolean operations at all but rely on statistical processing or concept-based searching, using techniques which can provide fairly good results but in which it can be particularly difficult to predict (or change) results. This is both because the exact algorithm of a particular implementation of a statistical technique is proprietary and because the processes are generally based on document structure rather than on rules of user input. In the absence of Boolean input rules, delimiters or document structure elements it can be very difficult to fine-tune such a search.

Finally, in the end-user world of the Web there is no tradition of exhaustive documentation for sources and structure, nor would such a thing presumably be welcome. But it doesn't take much experience with 'traditional' search systems to know that crude, unrefined searches of a collection of undocumented provenance is going to provide uneven retrieval for many kinds of searches.

2.1. Alta Vista

The phenomenon of a large ratio of false drops for every 'hit' is apparent in a system like Alta Vista, particularly if used in Simple mode as, presumably, it would be by most end-users. In Alta Vista there are two modes of operation, Simple and Advanced. In Simple mode there are no Boolean operators but there are techniques to indicate phrase and proximity, as well as to indicate prohibited words, and the system is basically a Boolean system with a classic inverted index.

Results in Simple mode are relevance-ranked, which makes up somewhat for the diffuse retrieval. In relevance ranking, documents which are statistically most likely to be relevant appear at the top of the document list. 'Relevance' is a statistical technique and there are many variations on how to calculate it. Generally speaking, the words in a query are compared against a document and the more words from the query that appear in the document, the higher its relevance is judged to be. Variations on this scheme include weighing words in a particular part of a document (say the title) more heavily; weighing words that occur often within a collection (such as 'company') less heavily; or ranking documents in which the query words that appear together, or nearby, are more highly ranked than documents in which the query words are widely scattered.

Relevance ranking can mean many things. In some respects what is really relevant to the user can only be known by the user. Much of what he or she would consider determinants of relevance, for example currency and provenance (article source and author affiliation), could be implemented simply as sorts or filters — but this would require more complexity of either the user interface or of the system itself. However when a search is extremely focused, such as a search for a known person's home page, relevance ranking in the statistical mode works pretty well. This kind of search pulled up several references to the page and its surrounding pages with the actual home page URL in the top ten of ranked items.

For a very focused search such as the one just described, Alta Vista has a major advantage — it indexes far more of the Web than its competitors and is thus the system most likely to locate a known item. It also has several advantages which commend it to experienced searchers. It has, for a Web system, fairly extensive Boolean capabilities and syntax in its Advanced mode. These capabilities include some field qualification, truncation and wildcards. Thus an experienced searcher can use nested Boolean logic to perform searches not unlike those on traditional systems.

As may be inevitable when so many sites are indexed, there are many duplicates retrieved with Alta Vista. And the Advanced mode general subject search, although 'good' compared to some of the other Boolean Web systems, is only about as 'good' as any search in a very large general database with no subject headings and no subject focus.

2.2. Yahoo!

Another way to structure information, long known to librarians, is to impose order on a collection in the form of an indexing scheme with an intellectually coherent outlook. At Yahoo! a staff of indexers determine how to classify each Web site. Each site goes into a regional hierarchy and is cross-linked to an appropriate topic. Thus a United Kingdom library would be listed under the *United Kingdom* with a cross-link from *libraries*. The classification 'scheme' developed by Yahoo! is unique to Yahoo! and is designed to at some point 'have captured the breadth of human knowledge.' This is a concept well-known to librarians and the limitations are also well-known: any attempt at ontology includes bias and for optimal use of a classification scheme, the user has to be familiar enough with it to think like the indexer — exactly what the prototypical 'end-user' is considered unwilling to do.

On the other hand, the power of a good classification scheme, or controlled vocabulary, is widely understood as a major advantage over keyword searching. A search on 'uk libraries' in Alta Vista retrieves a variety of library-related sites, most of them having nothing to do with the UK — on Yahoo! you see a selection of index terms which include *UK:Public:Libraries Page*, *Reference:Libraries:Countries:United Kingdom* and *Index — UK Higher Education and Research Libraries*, and links from any of these descriptors give right-on documents. Generally speaking, a Yahoo! search tends to have more precision than a search on a Boolean system. However, if all you

are looking for is a known item such as a company name, going through a controlled vocabulary is not the most efficient route. Also, if your topic is very precise then the chances that it will exactly match an index term diminish, as does the precision of retrieval. Finally, manual indexing is an inefficient, time-consuming process which tends to get behind the collection — especially if the collection is the size of the Web and growing exponentially.

In spite of Yahoo!'s superior indexing, a recent review of Web search engines in *Online* dismissed Yahoo! from consideration because the search engine was 'not powerful enough to handle our queries effectively' (Ref 2) — presumably in part because of an absence of fully-indexed documents with Boolean capabilities. And in fact, Yahoo! apparently realised these limitations because searches which do not get a 'hit' in the Yahoo! Index are automatically then run in Alta Vista. A search, for example, for a known home page brought up the Alta Vista results.

2.3. InfoSeek

InfoSeek is also an inverted index to Web Pages — however, once a simple index search is performed, documents are presented along with 'related topics' as well as a link from each retrieved document to 'Similar Pages.' A search for UK libraries on InfoSeek includes a rather odd array of related topics, from Law of the United Kingdom, to Novell NetWare, Maps and Veterinary Medicine. A 'Similar Pages' link from a document called *Web Resources in the UK* brings up a mixed bag of documents including *General Web Resources*, *Other World-wide resources* and *UK National Web and FTP Sites*. Generally speaking, one or two Similar Sites in InfoSeek may be of interest but the relevance is often marginal.

Although InfoSeek has no information on its own Web Page about it, it is reasonably certain that a statistical technique is being used to find related topics and similar pages. There are many variations on statistical analysis techniques but in a general way they work the same — a document or a collection of documents is scanned and statistics on the words in the document are used in various ways. For example, co-occurrence of words in a document or a document collection can be used to determine related words: if 'senator' and 'legislator' co-occur often then they might become related concepts. Of course, since computer analysis and not human intervention is being used, and since words or phrases can co-occur even when they are not related, then the results can be spotty.

There are many techniques to fine-tune automatic indexing and if statistical techniques are used, say, to summarise the content of the document then the statistical analysis may include a step which weighs terms differently, depending on where in the document they are located, or it will certainly weigh how frequently they occur in the collection against how frequently they occur in the document, for example:

'Keywords are typically chosen according to two standard measurements:

- (1) Words which appear frequently in a document are candidate keywords for that document.
- (2) The candidate keywords which occur infrequently in the rest of the collection are retained as keywords' (Ref 3).

The finding of Similar Pages could be based on something as simple as matching related concepts or be as fine-tuned as statistical clustering and categorisation — a technique in which documents are divided into subgroups according to their keywords and which is, in fact, the basis of the system surveyed below.

Generally speaking InfoSeek results in a subject search are interesting, if not exactly precise, and the system of related documents is quite good for browsing. At the time of this writing, InfoSeek is beta testing 'InfoSeek Ultra' which adds power with real-time updating, a number of semantic techniques (the system will find 'mice' if 'mouse' is input) and the ability to identify phrases, search exact case and so on. However, both of the InfoSeek systems failed to locate the known home page.

2.4. Excite

The Excite system uses the Architext search engine which indexes pages on the Web by concept. Although the details of implementation are proprietary, this is an example of the creation of a classification scheme through document clustering. Because document clustering techniques work from the actual collection, the subject categories are said to reflect 'the real world': if two words suddenly become related that were not related before, such as 'oj' and 'trial', then the subject categories will begin to reflect that immediately rather than having to wait for a human indexer to assign a new indexing term. And because clustering works on the entire collection, and because the collection in this case is large, then some ambiguities of meaning ('oj' as a celebrity or 'oj' as a citrus drink?) tend to be resolved by the clustering technique itself — if the user enters 'oj verdict' as a search then the documents that tend to cluster with 'oj and verdict' are likely all to concern concepts like 'trial, murder, sensational, century, lawyers' and so on, and not as likely to concern 'citrus, Florida, growers.'

For best results with a concept-based search, the ideal search will contain several concepts or a single unambiguous concept — a search on 'oj' alone doesn't gain any disambiguation by clustering. (However, if you want something on 'orange juice' on the Web it's best not to use 'oj' at all. The overwhelming results from 'oj' on any of the systems is OJ Simpson-related — for practical purposes, in the real world, 'oj' is now an unambiguous concept.)

Like InfoSeek, Excite offers a 'find similar' capability called 'query-by-example.' If an Excite hit is particularly good then the system will bring up similar items, relevance ranked (with the original good hit at the top). And if

all these tools aren't enough, Excite provides the user with the option to search by keyword or to use the statistically-created concepts. And all searches start as Boolean searches, using a 'fuzzy AND', which means that both AND and OR are used — with a higher weight being given to AND.

On Excite, the 'uk libraries' search retrieved, at the top of the list, items that were highly relevant — not as many relevant items as Yahoo! but more items, and more consistently relevant than the other systems. However a concept-based model, particularly one in which the concepts are generated strictly statistically, is still prone to errors and is stronger on recall than precision (although the known home page search did retrieve the home page and it was at the very top of the list!). It also can be difficult to determine why a particular item was retrieved and thus difficult to adjust one's search to correct problems.

Excite has a very nice display which includes a full abstract for most items.

4. Conclusion

The systems surveyed were good examples of the current state of Web searching. The InfoSeek Find Similar and Excite! concept search capability were not as precise as the human indexing of Yahoo! nor as imprecise as the 'simple mode' Boolean searching of an inverted index such as Alta Vista — but Alta Vista's 'advanced' Boolean did quite well also.

However, all of the current Web search systems are hampered by the sheer size and diversity of the Web, which makes it difficult to add value to 'documents' in the tradition of indexing and quality control, and also hampered by the Web's stateless nature which makes refining searches cumbersome and time-consuming. Depending on the search and the searcher, it might be better to opt for precision (Yahoo!); the large number of items indexed (Alta Vista); the browsing serendipity of related documents (InfoSeek); or the all-around performance of clustering techniques (Excite).

As end-users become more sophisticated about their needs, it seems likely that some Web search systems will begin to differentiate by topic areas, actively taking on more of the role of database producer, somewhat as Yahoo! is already doing with its indexing. It may also be that these very large systems are able to overcome the limitations of statelessness, much as some of the smaller, proprietary systems have already done, which would allow for delimiters and more ways to add value to the search process. I would predict that over time, Web search systems will come to resemble existing proprietary systems, and vice versa. The ideal searching environment, not easily realised either in twenty years of online or a few of the Web, combines a simple interface with sophisticated behind-the-scenes techniques. The Web, for the present, mandates a simple interface and there is a lot of experimentation with techniques.

Once the ease and power of statistical techniques are combined with the ability to refine user input, and reliance on the particular structure of documents or databases, then searching will get smarter — the goal is a system which works as well as an experienced librarian whose main business has always been to solve 'the information retrieval problem' and who has done it quite well overall.

Bette S. Brunelle
Ovid Technologies
333 7th Avenue
New York
NY 10001
USA
Tel: +1 (800) 950 2035, x 246
Fax: +1 (212) 563 3784 (fax)
E-mail: betteb@ovid.com

References

- [1] Steinberg, S.G. (1996) Seek and ye shall find (maybe), *Wired*, 4(05), 108-114.
- [2] Zorn, P., M. Emanoil and L. Marshall (1996) Advanced Web searching, tricks of the trade, *Online*, 20(3), 15-28.
- [3] Munson, J. and B. Thornburg (1996) Taking advantage of advanced searching, *Proceedings of the Seventeenth National Online Meeting*, New York, Information Today, Inc., pp. 263-268.



Online sources for competitor information

Arthur Weiss

AWARE, UK

Abstract: Competitor information gathering is a key aspect of business planning. Information will come either from published or unpublished sources. Unpublished information will often be verified based on material from published sources. Published information is more likely to be factual and includes financial, stockmarket, press, market and industry, product, trademark and patent sources. Much published information is now available online but the specific skills of a dedicated competitor analyst are required best to utilise these information sources. The traditional online hosts such as FT Profile, MAID, Data-Star and Dialog each host databases that can be used to obtain competitor intelligence. Typical databases will include those from Dun & Bradstreet, Reuters Textline, the Derwent World Patent Index, Kompas, Investext and others. The Internet has recently become an important source for competitor information, holding competitor promotional material as well as items from independent sources. The Internet search engines such as Digital's Alta Vista can locate information of interest on both the World Wide Web and Usenet.

Keywords: competitor intelligence, competitor analysis, online databases, World Wide Web, Internet

1. Introduction

Gathering information on competitors is a key aspect of business planning. Companies need information on competitors to compare their products, services and operations. The information gathered can be used to counter competitor actions and to set targets for company objectives.

Competitor information can be obtained from a variety of sources. These can be categorised as:

- unpublished information sources;
- published information sources.

Unpublished information sources tend to be internal to the company — the salesforce, other employees, suppliers, customers — or involve direct contacts with competitors and the industry, for example through exhibitions, market research studies and professional societies. Unpublished information is critical for gaining an understanding of competitors. It is more likely to give clues to an organisation's strategic thinking and tactical planning than published information. However a key problem with much unpublished information is that it can be difficult to assess its accuracy. Unpublished information may be misleading, false, or be based on rumours that will need to be substantiated. Often verification or otherwise will come from clues contained in published information sources.

Published information sources include annual reports and accounts, brokers reports, credit reports, press articles, directories, market research reports, patent information, company brochures, advertisements, press releases, conference papers and more. The number and variety of sources is numerous and growing.

Published information tends to be factual, verifiable data that at its most basic level can provide an understanding of competitor performance. It allows the calculation of market share and other comparisons between companies and their products. However it can also open windows into how competitors think, when an interview with a senior executive is published. Company promotions give a clue as to how the competitor sees itself and its market. Patent research can be crucial for identifying competitor product development plans.

Competitors can usually control what is published, ensuring that sensitive information is kept unpublished. However a skilled analyst can often gather clues to this information by collecting and collating data from a number of sources — looking at both published and unpublished information in tandem.

An added complication is the range of sources that can contain relevant information. To an extent, obtaining this information has become easier with the growth of online sources. However the increased accessibility of information has resulted in raised expectations.

These aspects of competitor intelligence gathering mean that a variety of tools are needed to gain sufficient information for effective business planning. Although the traditional information scientist or librarian will have the search skills for locating data, they will often not be familiar with the totality of available sources that can contain competitor information. Additionally, they will rarely have the strategic awareness needed to select those items

that can help the firm gain a competitive advantage over the other firms in the industry. At the same time the business planner will not usually have the search skills necessary to uncover the information. A third role — the competitor analyst — is required, combining the aptitudes of both these professionals.

2. Competitor analysis

Competitor analysis is a relatively new management discipline. Although the study of business competition has interested legislators and economists for many years, prescriptive advice helping businesses to succeed against competition is modern. For instance the leading management thinker, Drucker, writing in 1964 on how companies succeed virtually ignored competitive strategy as component of business success (Ref 1). Competitive strategy as a core element of strategic management has only really gained ground following Michael Porter's seminal works on the subject (Ref 2). The Society of Competitive Intelligence Professionals (SCIP) was only founded in 1986 — with 150 founder members. Since then the concept of competitor analysis as a separate business discipline has grown, with SCIP now claiming 3000 members working for companies ranging from AT&T to Zeneca Pharmaceuticals (Ref 3).

The skills required for competitor analysis are often confused with those needed for market research. This is a fallacy. Market research will generally be aimed at answering a specific business question. Competitor research is open ended. All information can potentially be useful. The good competitor analyst will have an innate curiosity, an ability to spot the seemingly inconsequential and the awareness needed to link this with other outwardly unrelated information to gain a complete picture.

The increasing availability and growth of online information sources has been a boon to the competitor analyst by allowing the analyst to monitor and search many more sources than had hitherto been possible. It has also resulted in the growth of a number of specialist management consultancies with the skills for searching a variety of online sources for information on competitors. Such services supplement the in-company competitor analysts, information scientists and strategic planners who may not possess the search skills or knowledge of all the databases available.

3. Online sources for competitor information

The range of online information sources available to the competitor analyst has grown considerably over the last decade. However with a few notable exceptions, the type of information available has remained constant.

Online information sources for competitor information can be categorised as follows:

- financial and credit information (for example company accounts);
- stockmarket information (for example broker reports);
- press information;
- market and industry information (for example market research reports and forecasts);
- product information;
- patent and trademark information;
- other information (for example conference attendees, legal information, contract and tender data, etc.).

3.1. Financial and credit information

Typical financial information will include company accounts and other financial statistics. Such information is readily available through a number of online services. The leading online hosts such as Data-Star, Dialog, MAID and FT Profile all contain financial databases from the leading company information providers.

Ideally, hard-copy annual reports should be obtained when examining corporate accounts. Printed annual reports usually include management profiles, pictures, outlines to strategy and notes to the accounts not held on most online databases. ICC produces a database containing the full text of quoted British companies, available via a number of hosts. This file can be useful for older UK reports but is expensive and excludes pictures and charts. An equivalent database is also available for the USA on Dialog (SEC Online — Annual Reports).

Understanding the financial detail in non-English annual reports can be difficult because of the different accounting standards as well as the language. Online databases such as Dun & Bradstreet's European Financial Records file on Data-Star, Dialog and FT Profile can help the analyst understand the accounts by providing an English translation. Dun & Bradstreet also has databases including basic financial data for companies in areas such as Eastern Europe and the CIS where it can be difficult to obtain any reliable company information. An added advantage offered by online databases such as the Dun & Bradstreet Financial Records file is the availability of financial ratios, which would need to be calculated if only the hard copy report were used.

As well as the online host services it is also possible to subscribe to company information services directly. Companies House offers an online company report ordering service for all UK limited company accounts (Companies House Direct). The leading business information suppliers also offer company financial information online — and in the case of Dun & Bradstreet and Infocheck-Equifax, this includes companies from most major countries worldwide. Proprietary services often offer additional elements not available through the online host

Online Information 96 Proceedings

services. For example, Dun & Bradstreet's D&B Access service provides a credit rating for each company as well as payment information. Credit ratings give an indication as to whether the competitor is financially sound while a poor payment score can indicate cash-flow problems.

3.2. Stockmarket information

A key information source for public companies is broker reports. In compiling such reports, market analysts will generally have access to company personnel. It is in the company's interest to be open with the analyst, to encourage a favourable report and buy recommendation. As a result, such reports often contain information that may not be available elsewhere — including strategic information and revenue forecasts. Broker reports are available on MAID, Data-Star, Dialog and FT Profile. For example ICC Stockbroker Research reports are available on Data-Star and FT Profile while Investtext Broker Reports are also available on Dialog.

Other stock market information sources can also be useful — Extel's information is available on a number of hosts and provides historical information on share prices and announcements, as well as financial information and extracts from annual reports.

3.3. Press information

Financial and stockmarket information sources provide extensive information on companies. However, perhaps the most important online information source for the competitor analyst is the press and news databases. The competent analyst will scan these sources on a regular basis, searching for any new news on competitor companies. Such information can be very varied and will depend on the type of news being reported — reports on competitor personnel, legal issues, new product launches, market sentiments, financial analyses and reports, and a whole raft of other issues that can help compile a picture of competitors will be published.

Press information may appear the easiest information source to use. However efficient use of press sources can be complex. A considerable amount of press information is repeated or out-of-date. The competitor analyst must be able to identify the types of information that will be useful while remaining cost effective. It can be easy to download every reference on a company but the search aim should be quality not quantity. Context display commands — for example the KWIC command on Data-Star — can help eliminate duplicate or irrelevant information. Appropriate use of keywords and search codes can also narrow the search. However items should not be eliminated based on length — even a single sentence news item can convey useful information. For example, typical press reports about job movers are only a few sentences long. Such items are useful in that they help in compiling a list of key personnel in a competitor, with their background and experience. If the position is new, the job placement may indicate a change in competitor strategy or organisational direction.

Not all press sources will be equally useful. It is worth considering what is included. One of the principle online databases for the competitor analyst will be Reuters Textline, which includes a large number of press sources worldwide ranging from the *Aberdeen Press and Journal* to the *Xinhua News Agency of China*. Textline also includes a number of specialist journals such as *Accountancy Age*, *Campaign*, *Estates Gazette* (useful for company relocation information), *Israel Business Today* and the *Vietnam Investment Review*. Nevertheless, restricting one's search strategy to Reuters Textline can result in the analyst missing a significant number of press reports. For example Responsive Database Services' Business & Industry Database on Data-Star and Dialog includes the UK information industry magazine *Information World Review*, absent from Reuters Textline. Using only Textline for a search on companies active in the online information industry could result in key developments being missed. Part of a search for news items should thus use databases that include industry specific journals. Other examples include the FT Energy Reports for companies active in fields involving energy or the environment, PIRA for the packaging industry and ABI/Inform for a number of the professions — accountancy, banking, insurance, etc.

3.4. Market and industry information

Market and industry reports help provide the context within which companies operate. Such reports often give an overview of the key players within an industry, including market share information. They also often include market forecasts and industry trends.

Market and industry reports will usually draw on similar sources to those used by the competitor analyst. As a result they may appear to duplicate information obtained from other, cheaper sources. However the reports can provide confirmation for this information and, being generally independent of the industry, give an objective viewpoint which may not be available elsewhere.

There are a number of providers of industry and market reports. They tend to concentrate on specific sectors or areas and no single provider covers all industries. For some niche areas, there may even be no good report available.

European market databases include Euromonitor Market Direction and Datamonitor, both concentrating on consumer sectors. Frost & Sullivan reports, ICC Key Note Market Analysis and MSI reports also cover industrial and business-to-business sectors. The latter two sources tend to focus on UK markets and there are also a few other country specific market trend databases (for example Databank's Market Structure and Trends in Italy). These databases are available through a number of online hosts Dialog, Data-Star, FT Profile and MAID, for example — although no single host carries all sources.

3.5. Product and trademark information

A number of databases provide brief information on the products supplied by companies. Such databases can be useful in identifying potential competitors in new markets, as well as possible suppliers to competitors (and potential suppliers to one's own company). Databases tend to focus on particular countries or regions — for example the 'Who Supplies What' database on Data-Star of German, Swiss, Austrian and Benelux companies and their products. A key online information source for product information data is the Kompass databases, available via the main online hosts and also directly from Reed Information Services' Reedbase Online service.

3.6. Patent information

Monitoring the filing of trademarks can give clues to new competitor product development. A new trademark from a competitor company — especially if it bears no relationship to an existing product name or is not an attempt to ring-fence corporate names to protect the competitor's interests — will often signify that a product with the trademarked name is about to be launched. Such clues to new product launches are an important aspect of competitor intelligence gathering, and allow strategies to be developed to counter the launch. Searches for new trademarks can be made via the Trademarkscan databases on Dialog and the Imsmarq service, also available from Knight-Ridder through a gateway.

For the technology, chemical and pharmaceutical industries patent monitoring is an essential aspect of competitive analysis. Patent filings give clues to the direction of competitor research and development activity. They give an early warning to the types of products being considered by the competitor which may not be ready for market launch for a period of several years. They can also allow an assessment of the strengths of the competitor's research and development team, by looking at the total numbers of patents filed by individual researchers in the competitor company, including patents they filed prior to joining the competitor.

Of all the areas of competitor analysis, patent information can be the most difficult to assess. Understanding the importance of a patent can require a high technical knowledge of the particular discipline covered. In many cases interpretation of such information is best done by the company's research specialists. However, using appropriate databases the competitor analyst can easily identify patents that could be of interest by printing the title and free search items of all patents filed by the target company. Patent searching can be restricted further by using the various subject and category search indices available with the databases.

The market leading supplier of patent information is Derwent, whose Derwent World Patents Index contains all patents filed since 1963, covering 40 patent issuing authorities. Derwent data includes the Derwent company code, inventor's name and a number of industry search codes. The Derwent World Patents Index database is available on Dialog, STN and Questel/Orbit. The information provided is an abstract of the patent, giving key details and where available an image of the particular invention or development. (Most databases miss out the images held in the original source. Derwent data is unusual as the images are retained. The only traditional host supplier that holds images on a regular basis is MAID, where a number of documents appear in the original source format, using Acrobat software.)

3.7. Other information

In addition to the above sources there are a number of other databases that can sometimes provide meaningful competitor intelligence. These will often be industry or issue specific. For example, the International Business Opportunities Service on Data-Star lists companies seeking joint-ventures and alliances with other organisations, while the Tenders Electronic Daily file gives details of invitations to tender for public supply and public works contracts. The UK Importers file on Data-Star lists all products imported into the UK, together with the names and addresses of the importing company, and Data-Star's ABC Europe: European Export Industry file lists details on exporters. Such files contain information that in conjunction with other competitor information could be important. A knowledge of corporate affiliations is also important — databases such as Who Owns Whom on Data-Star can give this information.

4. The Internet

Over the last two years, a key new information source has arisen as a result of the commercialisation of the Internet. Many companies now publish a considerable volume of promotional material on their World Wide Web home pages. Previously this information would only have been available through brochures which would sometimes be difficult to obtain directly. Further, the information published on corporate home pages is often fresher and more detailed than the corporate brochure which may only be produced annually. Some companies now publish their full annual reports on the Internet, allowing the analyst to cut and paste relevant details into spreadsheets for analysis. Significantly, some details may even appear before they are published in hard copy. Major companies also include the text of press releases on their sites. These may contain extra information not appearing in the versions that eventually are printed by the press. Hitherto, it was often difficult to obtain all a company's press releases.

The Internet is also providing other channels for obtaining competitor information — the majority at no cost. A number of leading newspapers now publish Internet versions of their paper, with full search facilities. UK examples include the *Times*, the *Daily Telegraph* and the *Financial Times*. The US patent office has made

available the full US patent file to searchers, giving patent information from the United States for free.

IBM's InfoMarket service (<http://www.informkt.ibm.com/>) offers the competitor analyst the opportunity to obtain information on companies from a number of databases, both free and charged. Databases available include those from American Business Information, Disclosure, Information Access Company's PROMPT service, MEDLINE and several more.

The InfoMarket Service is one of several search engines on the Internet. A typical search for information on a company may turn up dozens of references. For example the search results in Table 1 were obtained using two common search engines: Lycos (<http://www.lycos.com/>) and DEC's Alta Vista search engine (<http://www.altavista.digital.com/>). Although the results are for a very small sample, they show the differences between the two search engines. Lycos tends to be less comprehensive but always ranks the search results based on the likely relevance of articles to the search. Selection on the relevance score is possible, allowing the elimination of poor matches. Lycos also ranks each word separately — the exact number of hits for combined terms is not given.

Table 1: Search results for two Internet search engines.

Company	Search Engine	
	Lycos	Alta Vista
Amstrad	875	2000
Associated British Engineering	0	3
BASF	1230	4000
Boosey & Hawkes	171 with Boosey 1263 with Hawkes	600
Bula Resources	1	11
Lada	11	400
Matra-Hachette 1275 with Matra	732 with Hachette	400
Rentokil	40	189
Yorkshire Tyne Tees Television	10	117

A problem with searching for information on major companies is the potential volume of data available, especially if Usenet sources are included. A high proportion may prove to be irrelevant, especially as it is not always easy to limit the search to particular types of article. For example the Lycos search for Bula Resources only resulted in a reference to an academic paper where one of the authors was called Bula. In contrast to this, the Matra-Hachette and Yorkshire Tyne-Tees Television searches both listed annual reports among the first 10 items.

The search on Amstrad captured a number of Amstrad user opinions on the product, including one user asking for advice on a problem with Amstrad equipment. Such information can give clues to where a company has produced a product with design faults or other manufacturing problems. For this reason, restricting searches to the World Wide Web alone may miss potentially interesting information — Usenet newsgroup mailings should also be included within searches if available. It is not, however, worthwhile directly trawling through the various newsgroups looking for competitor information. Even though there are a few Usenet business sites for specific industries — for example *misc.industry.pulp-and-paper* or *misc.industry.utilities.electric* — the content rarely justifies a regular search. Relevant information is more likely to occur on an occasional basis in a number of news groups, picked up by the search engines as in the Amstrad example.

In addition to information on the Web there are a number of mailing lists for specific industries. These are often linked to specific Web pages found by searching for information on the chosen industry using search engines as described above. Some of these lists allow discussion between participants while others provide frequent news updates. Examples include a mailing list service from the Institution of Electrical Engineers (IEE), INSP-PRESS, giving news and press releases from the institution; a regular newsletter for the credit management industry which can be requested from <http://www.teleport.com/~richhl/>; and a weekly newsletter on the Internet and Consumer Online service developments (obtainable by sending an e-mail message to listserv@peach.ease.lsoft.com with <SUBSCRIBE ONLINE-L firstname surname> as the message text).

Arthur Weiss
AWARE
6a Northolt Gardens
Greenford
Middx. UB6 0LS
UK
E-mail: aware@dial.pipex.com

References

- [1] Drucker, P.F. (1964) *Managing For Results*, Heinemann, London.
- [2] Porter, M.E. (1980) *Competitive Strategy: Techniques for Analysing Industries and Competitors*, The Free Press, New York.
- [3] *Society of Competitive Intelligence Professionals Membership Directory* (1995) Alexandria, VA, USA



Creating services for the digital library

Dennis J. Crane

Corporate & International Markets, UMI, USA

Abstract: The terms 'virtual library', 'digital library' and 'electronic library' have received growing attention among professional librarians, researchers and users of information over the past decade. The confluence of exploding sources of data, expanding technical capability and constrained time and money will quickly move these concepts from exploration to execution by the turn of the century. Converting the potential of the digital library into reality demands a powerful array of services for users. Formulating and delivering those services will test many of the conventions under which researching and online services have operated over the past decade.

This paper proposes a focus on using software and electronic content to emulate interaction between the reference librarian and the library patron as the 'role model' to define and deliver highly valued 'Digital Researcher' services. These Digital Researcher services can fill a fundamental gap between the potential and the reality of digital libraries.

Keywords: Digital library, electronic library, natural language, services, UMI

1. Introduction

Some people go through life making things happen, some people watch things happen, and still others stand by and ask 'what happened?' In the world of libraries and information exchange, the confluence of technical possibility and human creativity have fundamentally altered the traditional view of the library. Many important changes are happening. These changes present great opportunity and challenge both for producers and consumers of information. These changes also alter the fundamental role of libraries in sharing information in a digital world.

As a newcomer to the field of secondary publishing and online database services, I would like to share my perspective on the transition underway in this market. I draw on more than a decade of applying information technology to change fundamentally the way organisations operate. I participated intimately in the somewhat parallel universe of business-to-business electronic commerce — the exchange of purchase orders, invoices and hundreds of other business documents from one computer application to another.

In 1984 Kenneth E. Dowlin, San Francisco's head librarian, authored a book entitled *The Electronic Library*. Slightly more than a decade later, the new San Francisco public library opened its doors and presented its patrons with a powerful example of a digitally enhanced library on a massive scale. San Francisco's library receives praise for being 'flexible enough to house computing technologies that are routinely outdated every 18 months, yet accessible enough to embody Dowlin's vision of the social role of the library' (Ref 1).

Technology has done much to capture, store, present and view a rapidly growing percentage of the world's information in digital form. Hundreds of specialised projects are advancing many aspects of the digital library. For instance, in the US the Networked Computer Science Technical Report Library (NCSTRL), pronounced 'ancestral', compiles leading edge computer research from around the world and permits Internet access. However, the project sponsors note that NCSTRL 'is not primarily about technology. It is about ... building the organisational and policy framework to support a digital library' (Ref 2).

From my perspective, the intelligent, informed, experienced reference librarian provides the greatest service value in many libraries. This implies that the service most essential to fully enabling the digital library will be 'Digital Researcher'. I would like to emphasise that the Digital Researcher will be a service, not just a technology. Some will react that technology cannot replace humans in this role. That's absolutely true. But computers, software and intelligent ContentBases can certainly and powerfully augment the traditional research professional. Technology will make many of the resources currently provided by the researcher more directly available to the user or patron.

By way of analogy, the ATM cannot entirely replace all the activities of the professionals in a branch bank. But, these devices perform many basic transactions quite reliably. More importantly ATMs provide new benefits, like 24-hour banking, that would be hard to implement in the traditional mode.

2. Understanding the digital library

We can envision combining technology and processes to create a Digital Researcher service capable of interacting with users of information to deliver valuable answers to their research needs. In defining the Digital Researcher we should place important emphasis on interaction. Striving to emulate the interaction that occurs naturally between the reference librarian and the patron has many valuable implications for defining compelling services.

Developers of software and online services, researchers in computational linguistics and proponents of artificial intelligence have pursued the holy grail of posing a question to an information system, in natural language, and receiving a highly relevant answer. This sounds exciting. It also ignores an important aspect of what makes traditional reference librarians and researchers so valuable. Interaction ... the seemingly simple ability of the researcher to relate with the user or patron through a series of questions, answers, options and choices to refine a query, and then deliver the most relevant response. Successful interaction goes well beyond composing a string of text and then using some Boolean operators to make it selective. A Digital Researcher must bring much more to the party. Drawing on a large and expanding intelligent ContentBase, and constantly updated knowledge domains, a Digital Researcher would add specificity and refinement through each direct interaction with the user population.

For example, user interest in the topic 'childhood factors influencing the character of Bill White' would elicit a response from the Digital Researcher service seeking to differentiate between Bill White the business executive, Bill White the sports figure or Bill White a local politician. The original query does not suggest anything about the profession of the Bill White in question but the user clearly has one of several individuals in mind. Why not use Digital Researcher's Knowledge Domain of people and associations to hone this definition, before spinning up the searchers to chase down the 'electronic stacks', retrieving articles of questionable relevance? Isn't this after all the manner in which a professional reference librarian or researcher would interact with the user to refine a query?

3. From data to answers

We are all aware of the explosion of data in digital format. IBM reports that the cost of digital storage has declined at 40% per year for the past 20 years and projects that trend will continue (Ref 3). Statistics on the World Wide Web indicate that popular Web search engines, like Digital's Alta Vista or Lycos, have indexed more than 20 million pages of content on the Web. This pales by comparison to the hundreds of millions of pages of digital content claimed by traditional online service and CD-ROM vendors like LEXIS/NEXIS, UMI, MAID and others.

This data explosion presents both a challenge and an opportunity for digital libraries and digital researchers. Claims of being 'the biggest' database have lost their meaning. The proliferation of Internet and Web access, plus the growing number of content licensing and reseller arrangements, act to make more of the world's published data part of one huge, loosely connected database.

The explosion of data in electronic form, however, places a much greater premium on finding the right information, in the right format, at the right time, at the right price to impact key decisions and actions. This motivates greater attention to understanding 'value in use' of any particular piece of information and the effort to synthesise information on any particular topic from all the right sources to provide answers.

4. From publication-centric to topic-centric

Traditionally, the creation and distribution of authored content revolved around publications. Publications — newspapers, journals, periodicals and potentially News groups and listservs on the Internet — provide convenience and value added through the article selection and editorial process. Information consumers gravitated to those publications that best served their needs. Logistics and costs of print and paper distribution limited the number of sources that a typical information buyer might utilise on a regular basis.

Personally I have chosen *Business Week*, *The Wall Street Journal* and *Fortune* as publications that keep me abreast of changes in the business world. I know that *The Economist*, *Forbes*, *The Financial Times*, *Harvard Business Review* and many other publications contain equally useful content and cover many of the same topics — I just don't have the time or money to subscribe to them all.

The advent of powerful computing, software and database services has fostered a subtle but fundamental shift in information access for busy professionals and researchers. Users increasingly value the ability to survey many relevant writings on a particular topic quickly and conveniently, regardless of source publication. Technology and the value added services of secondary publishers and database providers tap this customer need, which is inexorably shifting the pendulum toward customised, topic-centric information delivery.

The proliferation of current awareness and news filtering services attests to the appeal of topic-centric information delivery for business users, who are inundated with potential sources of information. Publications often expand their scope and content to attract new readers and advertisers. In the process they lose some of their specific relevance for some readers. They add 'overhead' to finding the information needed to support key decisions and actions.

Database search and retrieval have long concentrated on the topic-centric needs of librarians and

researchers. Indeed, libraries are the bastion of 'topic centred' information access — one location from which patrons could retrieve most of the information they needed to address a particular issue or question. With the advent of more user-oriented information services, libraries can now provide patrons with even greater capabilities and tools. These capabilities offer access to an attractive set of relevant responses, when and where user wants to receive them.

'Disintermediation' occurs in many industries and markets when technology augments or displaces the role of the 'middle-person' in the conduct of business or the exchange of value. Experience in banking, in many areas of retailing and in telecommunications suggests that the term 're-intermediation' is more appropriate. Technology and business process change combine to define new roles and opportunities for the service intermediary. Thus large retail specialists like Toys-R-Us or Sainsbury's provide a direct link between consumer and retailer. Technology displaces traditional inventory in the supply chain. Confronted with change, creative service intermediaries including librarians, researchers and secondary publishers must create new value propositions, often for newly defined customer segments, in order to prosper.

5. The Internet and World Wide Web

Re-intermediation, enabled by technology, will continue in the publishing and information delivery industry. The decade of the 1980s and early 1990s allowed an interested few to explore electronic and digital alternatives to traditional ways of publishing, accessing and exchanging information. The popular awareness and dramatic growth of the Internet and Web over the past five years have changed all that. According to Morgan Stanley, 'the market for Internet related products and services appears to be growing more rapidly than early emerging markets for print publishing, telephony, film, radio, recorded music, television and personal computers' (Ref 4).

An incident with my 11-year-old son personalised this fundamental shift earlier this year. Seeking to help him with a school research project, I suggested that we spend a Saturday morning at the library. He informed me, in all seriousness, that he did not intend to use the library for this project. Simply put, he said 'Dad, if it's not on the Web, it's not relevant!' While far from true, this sentiment reflects how pervasive the Internet and Web have become in the publishing and information exchange arena, and it's only just begun.

6. The Digital Researcher Imperative

The digital or electronic library promises electronically to capture and house vast collections of content, covering ever more sources from which topic oriented answers can be extracted ... but only if a 'digital researcher' exists and is up to the task of locating and extracting the right answers.

Herein lies the clear imperative for providers of services to the digital library — creating, operating, maintaining and enhancing a proficient Digital Researcher that intermediates between huge collections of electronic content and a growing number of technically aware and widely connected users.

Many consider a library to be a 'place'. Technology, communications and transitory lifestyles will make 'place' much less important. The ubiquitous cash machine emulates functions of the traditional branch bank but the technology, combined with human need and creativity, goes far beyond just replacing portions of the equivalent human activity. We would think it silly to limit ATM installations to three in a row, behind a counter at the branch bank, open only from 8 am to 3 pm (or whatever your country's retail banking hours). Indeed, we find ATMs in all kinds of places, from airports to hotels to food stores, operating 24 hours per day, seven days per week.

We should fully expect that technology, need and creativity will drive the Digital Researcher far beyond replacing some basic functions of a reference librarian and provide greater flexibility in the process. Research for this paper, for instance, was done entirely via UMI's ProQuest Direct from office, home and hotel rooms, in cities where it would have been inconvenient to visit a conventional library. The paper was then composed on a laptop, primarily on plane trips and on a Sunday afternoon by the community swimming pool.

7. Implications for service providers

The fundamental shift toward digital libraries creates considerable opportunities for new and expanded services for the proprietors of these libraries and for the users. Focusing squarely on the Digital Researcher concept can help vendors define and deliver services that will sustain value in the new information marketplace.

7.1. Creating and delivering value

Service providers must create and deliver compelling value if they hope to secure a position in the new millennium. Service providers must foster change and address the implications of that change. In this context, the definition of value and the means of delivering value shifts quickly to reflect underlying trends in technology, organisation and economics.

For instance, technology allows more primary publishers conveniently to create, deliver and archive an electronic version of their print publications. Witness new Web sites from nearly every major news and many periodical publishers over the past two years. Web publishing encourages inclusion of graphics and images that

are missing from traditional full-text databases. Many new Web sites have a 'nearly free' pricing structure, at least for now. These realities demand that database publishers quickly move beyond traditional roles of scanning, capturing and archiving text versions of publications. They must also expand beyond traditional indexing and abstracting to add more intellectual value to the aggregated content. They must synthesise answers from a range of sources.

7.2. Powerful Digital Researcher services

In this brave new world, successful Digital Researcher providers will combine key service features in a way the users will best describe as 'powerful.' The term 'powerful' includes, but more importantly integrates, many of the characteristics frequently ascribed to leading information services today. For instance, a powerful Digital Researcher service may not include the largest database but it must access content sufficient to deliver the best answers for the selected users.

Our customers are helping UMI understand that an appealing Digital Researcher service will combine the following key components:

7.3. Digital Researcher: intelligent ContentBase

The ContentBase for the Digital Researcher must compare favourably with the resources, contacts and collections at the disposal of an experienced reference librarian or researcher. The ContentBase must be vital and must incorporate new sources in a consistent and convenient manner.

Technology can substantially augment ContentBases in several ways. Computational linguistics allows analysis and updating of terms to help keep the value-added content fresh. Frequency, relationships and other statistics help refine information requests and improve response. Living 'knowledge domains' can classify key components of the ContentBase (people, companies, industries, products). Abstracting and indexing can associate specific pieces of content with topics of interest. Adding this kind of value to the ContentBase requires that the information be sourced, loaded and keyed consistently. This criteria will be hard to achieve in the rather random world of the Web. More importantly, all of this intelligent content can and must be linked in increasingly sophisticated ways to deliver real benefit to the user.

Intelligent ContentBases will take us beyond a traditional controlled vocabulary or thesaurus. Natural language and neural networking technology continue to make progress in capturing the essence of an article and mapping the content to other articles of similar nature and content. This is but one basis for linking items in the ContentBase. Another linking opportunity stems from the potential to tie relevant Web pages conveniently to companies, products, people, advertisements and other key concepts.

An article in January's *Chronicle of Higher Education* noted 'in recent years, many works have been translated into forms computers can understand either by taking a picture of each page, or by reproducing the text itself ... pictures give researchers more of a feel for the printed page and include illustrations, but doesn't allow them to search the text for keywords ...' (Ref 5). Today and in the future, this compromise should not be required. A truly intelligent ContentBase will contain items in varying formats, including a hybrid of full text for searching and links to high resolution images relevant graphic content. Imagine, for instance, the ability to link company or product names with scanned images of related advertising or product design.

7.4. Digital Researcher: interactive interface

A highly interactive interface must complement the intelligent ContentBase in order for the Digital Researcher to deliver valuable answers to users. Again, emulating the users' interaction with an experienced reference librarian or researcher provides key design guidance for this aspect of the service.

The Digital Researcher's interaction with the user combines associations in the ContentBase with response from the user to refine the selection, synthesis and presentation of the relevant answers. Counter to some views of natural language interface, the Digital Researcher would not seek a one-shot interpretation of the users' inquiry. Rather, the Digital Researcher would use several interactions with the knowledge accumulated in the ContentBase to iterate quickly toward the right answer.

Library patrons rely on the reference librarian to know the breadth and depth of an entire set of resources and to direct the patron's inquiry. Similarly, users of the Digital Researcher will expect the system to divulge facts and suggestions that refine a query through intelligent interaction. The Digital Researcher interface must conveniently translate user interests and needs into commands and instructions to the rest of the service.

Consider an early example from UMI's ProQuest Direct Search Assistant. Most users can understand the idea of constraining a search by date range, or by specific subjects or topics. However, few care to understand the complexities of Boolean search strings or remember the syntax for a specific online service. The Search Assistant feature interacts with the user, explaining options for focusing the inquiry. Behind the scenes, these interactions are converted to Boolean logic actually to direct the search. This form of interaction continues to expand to incorporate more of the intelligence in the ContentBase as the creativity of our application designers bring more sophistication to the software.

The Digital Researcher Interface can provide important dimensions of interaction. Users should for example be able conveniently to determine the cost to retrieve specific answers that meet their needs. Straightforward interaction with an ordering module, using graphics to illustrate cost of various options, can address this need.

8. Conclusion

The provocative term 'digital library' conjures up a vision of the electronic equivalent of an institution well known to all of us. Whether in a corporate, academic or public setting, a library has four key components: the building or facility, the content collection or holdings, the services of professional librarians and researchers and last, but not the least, the end-user or patron. Digital technology has already altered the first two components fundamentally. Physical buildings and facilities decline in importance as networks and digital repositories become more powerful and less expensive. More content can be found in digital form every day without ever needing to feel a thrill.

End-users or patrons will remain in non-digital form for the foreseeable future, some *Star Trek* characters notwithstanding.

This paper proposes that the missing link in the digital library remains compelling software and database services that can emulate and extend the role of the professional librarian and researcher efficiently.

Digital Researcher services will not displace information professionals, as some may fear. Rather they will extend and enhance the value delivered by these professionals. Modelled after the intelligent interactions which many in the audience have daily with your patrons and end-users, successful Digital Researcher Services will create value, answer questions and help expand knowledge well into the new millennium.

Dennis J. Crane
300 North Zeeb Road
Ann Arbor
MI 48103
USA
Tel: +1 (313) 761 4700
Fax: +1 (313) 973 9951
E-mail: dcrane@umi.com

References:

- [1] Markoff, J. (1996) Data on line and on shelves: libraries of future, books from past, *CyberTimes*, *The New York Times*, 15 April.
- [2] Anonymous (1996) Digital library system for computer scientists, *Information Today*.
- [3] Grimes, W. (1996) Libraries ponder role in the digital age, *The New York Times*, A.21, 29 April.
- [4] Meeker, M. and C. DePrey (1996) *The Internet Report*, Morgan Stanley, pp. 1-1 and 1-2.
- [5] Wilson, D.L. (1996) Language group urges book preservation in electronic era, *The Chronicle of Higher Education*, 12 January.



Preservation and access to manuscript collections of the Czech National Library

Vladimír Karen and Stanislav Psohlavec

Albertina icome Praha s.r.o., Czech Republic

Abstract: The Czech National Library started a large-scale digitisation of its extensive and invaluable collection of historical manuscripts and printed books in 1996. Each page of the document selected is scanned using a high-resolution digital camera in full-colour, processed and archived on a CD-R discs. HTML coded description is added to the entire document and to each page according to proposed recommendations currently under evaluation by the UNESCO Memory of the World Programme Subcommittee for Technology. Comfortable and easy-to-use Windows software, ManuFret, is available for searching and viewing the manuscripts at a higher level. The manuscript digitisation programme dates back to 1993 when the first pilot project was started and the 'Memory of the World CD-ROM' was published. Since then, several CD-ROMs have been published. Albertina icome Praha s.r.o. has been a partner of the Czech National Library since the very beginning of the programme and has actively participated in the preparation of the UNESCO recommendations. The collection of digitised documents is now open to the public in the National Library's manuscript study room and copies of the CD-ROMs are offered to other libraries world-wide on demand. The catalogue of the digitised documents is available on the Internet and Internet accessibility to full documents is being investigated. The referee will address technical aspects of manuscript digitisation and publishing that have been solved through the years, and software solutions will be demonstrated.

Keywords: digitisation, old manuscripts, history, CD-ROM, Internet, UNESCO

1. Memory of the World UNESCO programme

The Memory of the World Programme of the UNESCO General Information Programme dates back to 1992, when the first aims of the Programme were defined. The Programme, called also 'Safeguarding the National Cultural Heritage', aims to initiate, support and co-ordinate projects for preservation of cultural heritage worldwide. The basic idea is that a cultural society is obliged to collect documents about its own history, preserve them and keep them secure for future generations.

Documents of a great historical and cultural value are often requested for studying purposes. Removing them from optimised climatic conditions in the repository can, however, cause them irreparable damage. Even slight, hardly noticeable damage can result in the loss of such a document. There are numerous examples when even a careful, qualified restoration process turned into a disaster. The primary task is therefore to prevent usage of the original documents. The secondary task is, on the other hand, making the documents accessible by providing acceptable replacements using a state-of-the-art technology.

2. History of the Czech Project

The Czech National Library, along with the private company Albertina icome, were among the very first participants in the Programme and a contract with UNESCO was signed about the development and publishing of a pilot Memory of the World CD-ROM at the beginning of 1993. This CD-ROM contained bibliographic descriptions of 150 of the most valuable manuscripts and old printed books from the extensive collections of the National Library. Along with bibliographic records, over 100 digitised photographs of these documents were included. The entire task should have been completed in only four months from the final decision to a published CD-ROM discs! This was an extreme task but the joint efforts of the National Library and its private partner proved successful, and the product was presented at the UNESCO meeting in Warsaw in May 1993. This was the very first outcome of the UNESCO Programme. It helped to define further tasks and started a discussion on technological aspects of digitisation.

The following two years provided time necessary for discussions and gaining expertise. Finally, the first

Online Information 96 Proceedings

Page 403

manuscripts to be digitised cover-to-cover appeared in April 1995. The *Antiphonarium Sedlecense* (Antiphonary of Sedlec), a unique illuminated 13th Century songbook, was selected and digitised using a high resolution digital camera. In addition to ordinary bibliographic information, detailed descriptions were created by a team of experts from the point of view of a historian, musical historian and a codicologist. At the same time, another digitised manuscript — the *Chronicon Concilii Constantiniensis* (Chronicle of the Council in Constance) — was prepared thanks to support from the Foundation of the National Library and its private sponsors. This was the pictorial addendum to the official chronicle of the Council and it was created by Ulrich Riechental in the first half of the 15th Century. This document is quite unusual in showing not only the church officials and the council alone but also the life of ordinary people at that time. Despite its mainly pictorial contents, the full text of the manuscript in Latin was included on the CD-ROM along with its translation into modern Czech. These two products were the first of the *Memoriae Mundi: Series Bohemica*, a series of digital facsimiles of Czech cultural heritage.

As well as the project in the National Library, projects with other institutions were also realised. Collaboration with the State Central Archive in Prague resulted in 1994 in the electronic edition of a full collection of correspondence from the pen of Clemens Lothar Metternich called *Chancellor Metternich and His Times*. Metternich is known as a primary Austrian politician and he played a major role in international development in the first half of the 19th Century. He was in touch with many important personalities of his era such as Napoleon Bonaparte or Charles Ligne. The CD-ROM contains nearly 2000 digitised pages plus an extensive database of other material. Another project with the Museum of Czech Music ended in the full digitising in early 1996 of *The Two Widows*, an opera by a Czech composer, Bedrich Smetana.

3. Preservation issues

There were plenty of questions to be solved during the past years. One of the very basic concerns was to minimise the need to make the original documents accessible to experts for studying purposes. However, the question was 'will experts really be able to live with the electronic facsimiles instead of the original manuscript'? Numerous surveys and tests showed that (after a necessary 'acceptance' period and a minimum training) a minimum 95% of the experts were able to fulfil their research needs by using the electronic document. In fact, all of those interested in the pure contents of the manuscript were satisfied — moreover, usage of the electronic document even lead to discoveries of certain aspects that were never seen in the original work! In addition to that, the experts appreciate the removal of the stress factor experienced previously when they were only allowed to work with the manuscript for a strictly limited time. The only ones unsatisfied were those studying the material aspects like the writing material, inks and colours used etc. — but they will never be satisfied by any replacement if not 100% identical to the original. Anyway, it was possible to lessen the unwanted 'load' on the originals caused by the need for making them available for top researchers to as little as 5%.

4. Access issues

Although preservation is the primary concern in the project, the access issues appear to be even more exciting. When speaking about experts ever allowed to touch a manuscript we are talking about maximum of a dozen of the topmost specialists annually, sometimes even much less. However, the 'electronic manuscript' can be made available not only to specialists but also to students of history and virtually to anyone interested without any danger to the original. This will again completely change the manner of education in this field as the students can be allowed the same opportunities to access the historical documents as their patrons — even though some of the latter will probably not be very happy about this idea.

However, other questions have arisen as more of the Memory of the World projects started to appear worldwide. Probably the most important one was precisely because of the number of projects — as each of the projects used slightly different approaches, they used different media and various software. As a result of this, there are several products in completely different data formats working on different computer platforms. However, preservation should also mean that the new electronic format should even survive the original document. In this case it would not make sense even if the CD-ROM would last hundreds of years, when in just a dozen years there may not be a computer system available to work with the data, not to mention the software part of today's products. A solution to that can only be in adopting a published and well-known data format common for all of the future projects in every library around the globe.

As far as the image format goes, the above concerns lead to a UNESCO recommendation to use JPEG compressed image files for the Memory of the World projects, because of the excellent quality to file size ratio of the JPEG format. Also, JPEG files can be displayed at a reasonable speed and are compatible across a wide range of computer platforms. Again, our surveys showed that it is not possible to define a 'minimum' or 'recommended' image resolution. This is because the necessary resolution is strongly dependent on the document itself (its quality, size, conditions etc.) and can vary over a large range (Ref 1).

However, the digital facsimiles (or digitised books) do not consist only of images. They should also contain descriptive textual information, from a simple pagination to the full text of the document. This meant that we cannot take advantage of common relational databases because they cannot handle full text. On the other hand there are already standards in the library sector defining the contents of such textual data (e.g. AACR2). Following extensive research in this field, the team of the National Library and Albertina i come Praha suggested a common

Online Information 96 Proceedings

Page 404

defined format based on HTML (hypertext markup language) (Figure 1). Within HTML the format defines certain tags that can be used to create a full-text database structure that can be used while searching the database. This format was proposed to the Subcommittee for Technology of the Memory of the World Programme at its meeting in Prague in March 1996 and is currently under evaluation. It is highly probable that this proposal can be accepted as a UNESCO recommendation for the Memory of the World Programme in summer 1996 (Refs 2, 3).

The acceptance of the descriptive format proposal would mean another qualitative step in the Memory of the World. This would imply that all the data created within the Memory of the World will be fully interchangeable and both media and platform independent. Such data can be accessed at the minimum level by any HTML (i.e. World Wide Web) compatible browser like Netscape, Microsoft Internet Explorer, Mosaic etc. However, it is predictable that more specialised software packages will be developed to provide researchers with much more functionality needed for searching in the documents at high speed, fast retrieval of images, image evaluation and so on. As an example, Albertina i come Praha developed *ManuFret*, a specialised manuscript-oriented software package based on our proprietary full-text technology called Fret (Figure 2). *ManuFret* can input HTML coded data automatically, index them and provide all the necessary features in a very simple to understand environment. *ManuFret* is available free to all users of digital facsimiles created by Albertina i come Praha within the Memoriae Mundi Series Bohemica project, as well as being part of a knowledge transfer programme to other libraries worldwide.

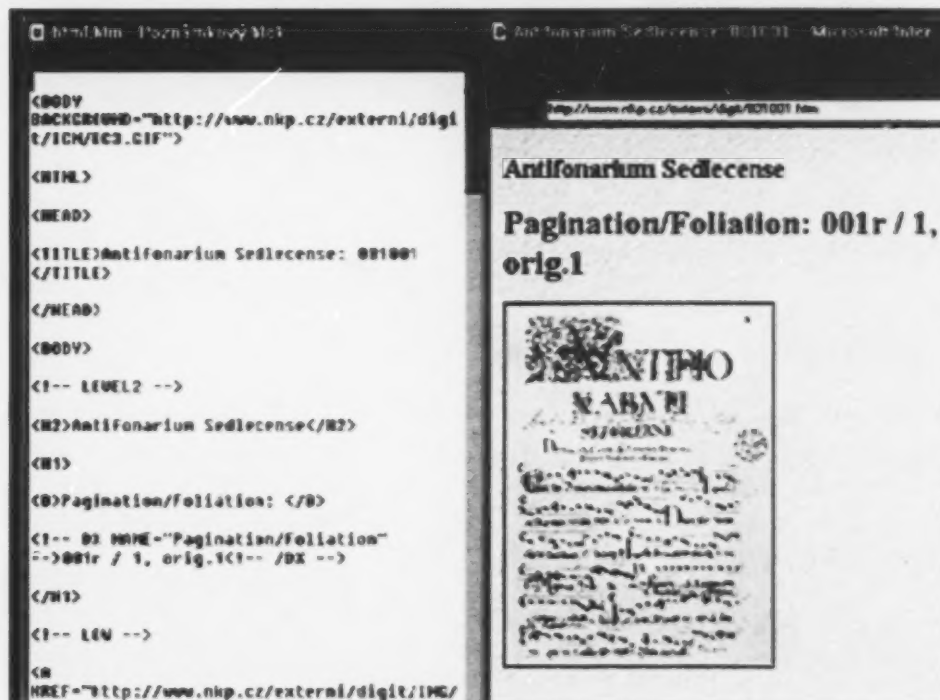


Figure 1: Example of a manuscript's page (Antiphonarium Sedlecense) description in the proposed format: HTML code (left) and its representation in a WWW browser.

are also examining the possibility of providing access to digital facsimiles in lower-level quality via the Internet, since the bandwidth is not yet satisfactory for high quality (i.e. large volume) images online. Full quality digital facsimiles on CD-ROM/CD-R are available to other libraries or individuals on demand within the framework of the Memoriae Mundi Series Bohemica Subscription Programme, or on a one-time purchase basis (Ref 2).

Another task will be the establishment of the Memoriae Mundi Series Bohemica Digital Archive. Its aim is to maintain the collection of digital facsimiles created within the project. The Digital Archive will be responsible for storage of all the digital data and, eventually, for their conversion into any future digital format. A high-capacity storage solution will be needed in the more distant future once the decision to go online with the collection has been made.

There are many opportunities for foreign collaboration. Our project is open to any reasonable collaboration and we are ready to help other libraries in starting similar projects with our experience, know-how and software. For example, one such opportunity can hopefully start in collaboration with the Austrian Ministry of Education and the University Library in Graz, Austria.

7. Conclusions

Digitisation of old manuscripts and printed books is coming of age in the Czech National Library. We can foresee years and years of work in order to fulfil the goals of our project. On the other hand it is very exciting and enjoyable to see the project growing from the very beginning, and we are very proud to be at the cutting edge of developing a small but important part of the Global Village.

What would it mean if your library decides to start digitising its entire manuscripts collections?

You would need to:

- (1) decide whether your collection is large enough to justify the costs involved in building your own system, or you should better join forces with others;
- (2) contact UNESCO GIP to be listed in the Memory of the World Programme;
- (3) find the funding for (maybe) a couple of years of work;
- (4) learn about existing or *de facto* standards;
- (5) test the equipment in order to achieve proper set-up and personal experience;
- (6) and then go!

We will be more than happy to help you with any of the above issues.

Vladimir Karen
Albertina icome Praha s.r.o.
Revoluční 13
110 00 Prague 1
Czech Republic
Tel. +42 2 2480 3303
Fax: +42 2 2480 3296
E-mail: aip@login.cz

Stanislav Psohlavec,
Albertina icome Praha s.r.o.
Na drázkách 328
266 01 Beroun 2
Czech Republic
Tel./fax: +42 311 621053
E-mail: aipdev@login.cz

References

- [1] Knoll, A. and S. Psohlavec (1995) *Digitization of Old Manuscripts*, Czech National Library, Prague.
- [2] Albertina icome Praha WWW home page — <http://www.aip.cz>
- [3] Czech National Library WWW home page — <http://www.nkp.cz>



Down to earth in cyberspace. Building an electronic library: policy, plans, practice and problems

Natalia Grygierczyk

Utrecht University, The Netherlands

Keywords: electronic library, guiding policy, project management, users' requirements, information retrieval, electronic collections, infrastructures

1. Introduction

'Suppose that an ephemeron (a mayfly), hovering over a pool for its one April day of life, were capable of observing the fry of the frog in the waters below. In its aged afternoon, having seen no change upon them for such a long time, it would be little qualified to conceive that the external branchiae (gills) of these creatures were to decay, and be replaced by internal lungs, that feet were to be developed, the tail erased, and the animal then to become a denizen of the land.' Thus wrote the Scottish evolutionist Robert Chambers in 1844 (Ref 1).

As a metaphor Chambers' assertion also applies to the way we humans observe several phenomena. Yet there is an essential difference. The moment an animal dies its experience and knowledge vanishes forever, with the exception of a fraction that is transferred genetically and takes thousands of years of evolutionary adaptation to change.

We humans, though, are capable of preserving the knowledge we acquire and transferring it without the necessity to encode it genetically. The independent transference of information enables us to undergo a cultural and scientific evolution proceeding at an incomparably higher pace than any organic evolutionary process.

The provision of information through electronic networks marks in fact a revolutionary stage in history, though of course we're not busy pondering this thought all the time. Not even you and I, who as participants of this conference are daily engrossed in the problems of electronic information.

'The post-information age is about acquaintance over time: machines understanding individuals with the same degree of subtlety (or more than) we can expect from other human beings, including idiosyncrasies (like always wearing a blue-striped shirt) and totally random events, good and bad, in the unfolding narrative of our lives. [...] Like a force of nature, the digital age cannot be denied or stopped. It has four very powerful qualities that will result in its ultimate triumph: decentralising, globalising, harmonising and empowering.' Dixit Nicolas Negroponte (Ref 2).

It's quite obvious that Internet and the computer are capable of evoking a plethora of prophetic visions as to where our world is heading. Every expert and prodigy of learning is cherishing his own opinion on the matter. In the debate on the Internet two opposite camps are engaged in a fierce battle.

The first faction consists mainly of hosanna-criers who immediately get carried away by the word itself. 'It suffices to type down one term' — they preach us — 'and the whole world shall be yours on your screen.' Splendid indeed. Yet who's craving for this?

The second party — slightly less enthusiastic — are foreseeing a disastrous mental pauperisation. In the near future — they're predicting — we needn't remember anything ourselves anymore. We must only be able to locate the right home page. Before very long the structure of knowledge will be changing drastically owing to the conviction that we can safely rely on an externalised memory. Because of the hyperlinks we're going to lose the ability to follow a linear course of reasoning, thus limiting ourselves to a chaotic criss-cross way of thinking. We shall unlearn the art of deducing. Eventually, we are warned, the majority of society will consist of infosurfers, socially and emotionally handicapped beings net surfing their way through life. Beware ye all, the creation of the 'homo sapiens' is at hand.

'Down to earth in cyberspace' is the title of the present paper on the making of an electronic scientific library.

To us, builders of such a library, the Internet, computers and electronic information sources are no more than useful, perhaps very useful means that can be employed efficiently in order to improve what has always been important first and foremost: scientific education and research.

Within this context a certain reticence may prove quite commendable. I'm afraid I therefore cannot present you with some terribly exuberant ideas.

Scientific knowledge possesses structure not because it has been documented in one way or another — on paper or in cyberspace — but because we've been giving our thoughts a structured shape. The Internet in the first place should be equivalent to access and communication, so it hasn't got much in common with the structure or the content of scientific knowledge. Stated otherwise: the meaning is *not* the message.

2. Electronic information policy: why?

Internet, electronic information provision, multimedia; in the applications for subventions and grants they're used as magic words. But don't worry: I shall not be discussing the present reallocation of means from a traditional to an electronic provision of information. The growth of electronic collections, the disproportional rise in prices of licences for the use of electronic information, the cancelling of subscriptions to paper versions of magazines because of financial cut-downs — all this has already been discussed sufficiently at university libraries. I'd rather pose some questions that are rarely asked when we're dealing with projects concerning the electronic provision of information.

Why is this sort of electronic provision useful? Do we really need it? Who's going to use it and for what purpose? Are there any existent examples? Are all the possibilities of the medium extensively exploited or does it merely boil down to a copy of a manuscript in an electronic form? What about the quality of online files now and at sometime later? Who takes care of the quality? How is it guaranteed?

All these questions should be asked and properly answered in any project dealing with the electronic provision of information. Better even if not only the contents of an individual project were scrupulously defined but also the framework of its subsidiary projects.

It proves there are many projects at universities and other institutes that have been approached and taken up arbitrarily just because there happened to be a driving force or some financial resources as a result of 'grant chasing'. Quite often these are incoherent projects lacking a well-defined policy. In some instances there is even hardly a necessity for any project result at all. It also occurs that similar projects are executed elsewhere — sometimes at the same university. There are also projects the outcome of which vanishes forever in some desk drawer just because the project management fails to take in account the need to incorporate the results into the regular organisation.

Therefore I'd like to describe the approach of a project initiated at Utrecht University. In our view this approach contains a certain guarantee to avoid a few recurring problems and pitfalls. It is based on a coherent project organisation and has got a clear project framework from which several subsidiary projects have been derived directly.

I'd also like to say something about a number of practical problems and their solutions, as well as the human factors within the project management. Thus, a short case story.

3. The electronic library at Utrecht University: points of departure

The Electronic Library Utrecht is a project at Utrecht University, a large comprehensive university with 25,000 students, 7000 employees, 14 faculties, 19 research schools and a library collection of 4.5 million items. The project is a joint initiative of the University Library and the Academic Computer Centre Utrecht. The project was initiated in 1995, has a duration period of three years and has a total budget of £850,000. In fact it consists of 40 systematically interdependent projects. All these subsidiary projects have one common goal: the qualitative and quantitative improvement of the present provision of information, a better accessibility of information sources, and last but not least the development of a series of completely new library services.

The leading principle of the Electronic Library Project is to comply with the requirements and wishes of researchers, teachers and students. In this respect the library project is not so much a technology-driven project (based on what is technically possible) as it is a user-driven one. This implies that our users are continually involved in assessing both purpose and result.

How to take up such a series of projects? What should be its guiding ideas, its starting points? How to organise the project itself? Who is to participate in it or to contribute to it? To us all these questions had to be answered referring to the needs of the users.

3.1. Users' requirements

Yet how does one find out what wishes and needs all these users have? Just by asking them, one would suggest. However, the question of what library services they want to see realised is not very likely to yield really useful conclusions, for researchers and students simply haven't got an understanding of every alternative, nor of present and future technical solutions.

Hence the user survey at the outset of the Electronic Library project aimed at tracing the *problems* students and staff were having with the electronic provision of information. Its second purpose was to map the common *practice* of scientific research and education, i.e. how do you start preparing an article, a paper, a new series of lectures; what sources are used; and how do you search for them (Ref 3).

The results of the survey were analysed: users' requirements were derived and translated into several facilities feasible in the short and the long term, thereby taking into account the present state of technology, future standards and foreseeable developments. From the analysis have accrued a number of technologically sophisticated projects as well as some relatively simple, even trivial solutions for infrastructural problems.

The conclusions and the different project types are described in the next section. First, however, I'd like to mention our remaining points of departure.

3.2. The integration of physical and electronic collections

Suppose you were a physicist sitting at your desk. You'd like to have a look at the latest research reports on magic numbers in helium clusters. So you start up your computer and ask a database your question. It gives you a list of papers in the *Chemical Physics Letter* plus some book titles. You read the abstracts on your computer screen. One of them may be very significant and you switch to the full-text version.

Later you have another glance at the titles. One treatise looks of interest to you: it's on classical rare-gas clusters. However, the treatise is not available in an electronic form. So you switch to the library's catalogue. The book appears to be in your own library's collection. The only thing you have to do now is to order it online: within an hour the book will be right on your desk.

In many libraries this is of course still a fancy, though not for much longer. This example illustrates the thought that an electronic library isn't an isolated, virtual building in no way related to a physical collection — a collection that has been gathered, administrated and disclosed since the times of Alexandria.

Electronic and paper collections must be integrated in a complementary way. This objective implies an inter-connection between primary and secondary information; between on the one hand full-text files or multimedial electronic files and on the other bibliographic sources that refer to actual books and magazines, and their physical location.

For the user this principle should bring about the possibility of looking for information on a certain subject without having to indicate in advance a source or collection. This brings us to the next principle.

3.3. Searching for information independent of technical structures

Let's first return to reality. Suppose you were engaged in medical research on spontaneously epileptic rats. Suddenly you have an urgent question, so you turn to a database containing abstracts from a magazine called *Brain Research*. You look into the next database, one that is not linked up with the first. But it appears you're not very familiar with the latter type. It takes you quite a while before you understand the system and the function of several keys and buttons. You're formulating and reformulating your question over and over again. You start searching in the catalogue. Quite desperate, finally you turn to Internet. Incidentally, you glance at your watch and breathe a deep sigh.

Can we apply information technology and all its promises, avoiding such time-consuming quests for information itself? We think so, for instance by designing an information retrieval system that enables you to browse several databases (if required) by means of one search question. This system should operate on bibliographic files as well as full-text files and other sources. It should be able to go through files marked with different indexing techniques and it's got to cope with a dynamic, physically distributed and ever-increasing collection. In short, an information retrieval system set into motion by a single entry question whilst sparing you the technical idiosyncrasies of certain databases.

3.4. An information retrieval system dependent on the type of user and usage

There are still some more general demands to be made to an information retrieval system. The system ought to operate in a tailor-made way, i.e. in accordance with the needs and wishes of the user.

Suppose you were a barrister as well as your job at university. You're in a hurry because you've got to attend to a session in court in two hours. You know there is quite some jurisprudence in the case you are defending today. You also know where to find the information. But you don't want all of it and certainly not any theoretical data. You just want a specific answer to a specific question.

Suppose, after a successful session, you sense the inspiration to write a paper on the matter. In this instance you'd like to gather as much theoretical background information as possible: not only legal information but also all the relevant medical data you require. Now your wishes are of an entirely different nature to those of a few hours before.

This example suggests that an information retrieval system suiting the demands of every university staff member and student must also have a great deal of flexibility. It should be able to serve many different users who may have many varying needs.

In accordance with this the system interface should be flexible too. As a lawyer you do not wish to analyse the mechanics of the system. Nor do you feel the urge to study the inscrutable formulae in which your search question should be cast. No. You prefer the use of vernacular language using perhaps a few specific legal expressions. However, put your modest request to a forum of reference librarians and they'll overwhelm you once again with an avalanche of Boolean formulae, fascinating in their logical elegance, though quite cryptic to you.

3.5. The digitalisation of the library's own collection amounts to the digitalisation of the entire service involved

Finally the last principle. When we're discussing the building of an electronic library and the digitalisation of the library's physical collection we don't want to limit ourselves to the digitalisation of the information sources.

Even in a very modest case when we're dealing with the digitalisation of a small part of the collection, we should digitalise the library service itself. For crucial to the availability of electronic information is the adjustment of the library structure and organisation round this source. Consider for example the difference in maintenance between an electronic information source and a book or paper magazine: think for instance of the struggle against computer viruses versus moulds devouring paper books.

Less obvious, however, is the implication that library staff see their functions redefined. New skills are being demanded from them. Users will be calling on them in a different way. Consider for instance the gathering and administering of new and old material. One would expect the library organisation to apply the same quality demands to electronic information as to all other sources.

This strictness is especially significant in handling online files on the Internet. Clear agreements must be made with suppliers and manufacturers on the sources' guaranteed unfaltering quality. As a reader one wouldn't like reading an article suddenly missing a few pages that have been changed overnight.

4. Project organisation, communication structures, implementation of results

By now you may be thinking, how thoughtful to take up a project so nicely by the methodological book, neatly defining all points of departure and elaborating a long term policy. Yet, somehow and somewhere we've heard the story told before. At similar conferences we've been witnessing plenty of presentations, all of them quite impressive in their elaborate structure. So please allow us to ask, what is actually realised?

I cannot but agree with your scepticism. Nonetheless, before I answer your question, hoping to convince you by pointing out some actual results, I'll describe the guarantees built into the project organisation in order to limit the risk of failure to smallest possible degree.

First, we're dealing with a university. Utrecht University is characterised by a large degree of decentralisation. This means that every department and faculty is relatively independent. These faculties work out their own policy, not only on education and research but on automation as well.

Of course, this decentralised structure offers many advantages. The University Board can concentrate on the main management issues without being distracted by details. From a strategic point of view it can lead to quality improvement of decision-making. The departments are responsible for their own resources. Thus the resources and tasks are interconnected at a level where the consequences of the policy can best be overviewed.

For us, alas, this structure has some disadvantages. Providing facilities to research and education, this project may not easily count on a smooth acceptance and use of its results. The only method (Greek for 'way about') is the method of consensus.

Important therefore is the existence of so-called communication structures between the project organisation and the departments' managing directors and automation experts. A significant factor in this constellation are the contacts with the eventual users, the students and staff of the department. The department's board usually lends a more attentive ear to the interests and wishes of its own users than to those of managers belonging to a project that isn't part and parcel of the faculty.

Another measure to safeguard the outcome of the project is of an organisational nature. Many IT projects find their apotheosis in the approval of a thick report containing lots of recommendations, schedules and diagrams. Thereupon the report is often put into safe keeping for the next decade or so.

To avoid such an unfortunate fate the results of each subsidiary project of the Electronic Library Utrecht are implemented in the regular organisation and its branches. This implementation is conducted in the last stage of each project. The altered tasks of library staff and Academic Computing Centre personnel with respect to the new services are first described and then introduced at each department. Obviously, this is not possible without the unconditional support of the university librarian and the managing director of the computing centre. To that purpose both directors have been appointed as supervisors of the project and of the implementation of results in the steering group.

5. The realisation of the projects

What is actually being realised? What is the content, the body of the projects? Do they really comply with the guiding principles?

I've already mentioned the point that the user surveys, in which *problems* were assessed and analysed, provided a guiding factor in the definition of our projects. One can easily guess that a larger part of these problems was connected with the infrastructure and the support during information searches. The first series of subsidiary projects is dealing with the infrastructure in its most comprehensive meaning.

The same user surveys were used to analyse the *wishes* of students and staff as to library services. In addition the project management of course took into account a broad range of technical facilities, new standards and trends. These requirements, now in specified form, have led to a second series of projects, i.e. the new electronic library services.

5.1. Project series 'Infrastructure'

One of the most astonishing conclusions of the user surveys and panel debates conducted amongst users in Utrecht and at universities abroad was the assessment of a stark need for (electronic) information sources, whereas at the same time — and this is the astonishing fact — hardly anyone knew of the availability of already existing facilities. If someone did know he wouldn't consider using them.

Why? one is tempted to ask. The simple answer is that all these information sources aren't accessible when

you are sitting at your own desk. Students and staff have to go to the main library building where they are welcomed by a computer screen in some dark corner, showing an eery interface with grinning codes. Even for people with steel nerves this proves too much. The timid scholar seeks safety with a library employee who appears to be extremely busy. So in the end he runs off back to his refuge. And there he's at his desk again patiently waiting for the information in print they've promised to send him by the end of the week — the same information he could easily have had on his own computer screen within a few seconds.

Lately, the price of information has been increasing sharply: not only in an absolute sense but also in comparison with the frequency it has been used, as is the case with electronic information. A peculiar situation, for more information is being offered (at a steadily growing price) whereas users tend to neglect it. Yet this regrettable — some would say unacceptable — fact is a general phenomenon at many universities and institutions.

The solution is self-evident: better information, guidance and instructions on the availability of electronic information sources and its facilities, improved ergonomic interfaces, an efficient information retrieval system, the additional training of library personnel and an uncomplicated access to all information sources. These objectives are being realised in subsidiary projects. Each of these projects has a guideline defining its substance in accordance with the leading principles of the project's framework.

On the one hand the projects aim at solving existing problems. On the other hand these solutions must be characterised by definite durability, taking into account the dynamic developments in the field, for example an ever increasing collection, changes introduced by information suppliers who tend to offer more and more platform-independent sources (like SilverPlatter), the expansion and improvement of users' equipment, the library's independent managing and maintaining of electronic information sources — several professional consequences for the library staff.

In the section after the next I hope to give you an idea of these kind of problems and solutions in practice.

5.2. Project series 'Electronic Library Services'

The leading principles for this series of projects were also derived from the user surveys. In this instance we've concentrated on the wishes expressed. The projects include:

- the digitalisation of traditional services:
 - the filtering of information and electronic notifying services (SDI);
 - the electronic availability of the books' table of contents;
- the digitalisation of the library's own collection: the (partial) digitalisation of special collections, old prints, impressions and manuscripts;
- the digitalisation of the university's scientific output: electronic publishing of theses, papers and articles.

In this case we also keep ourselves to a number of guidelines. For instance a digital information source ought not be a mere electronic copy of a paper one. Besides linear scrolling it should be possible to indicate a desired topic. Apart from text, pictures and statistics, one would like to include moving images (films) and sound. Separate documents should be hyperlinked to related material within the library's own collection and beyond, but also indicating bibliographical sources and the location of printed material.

These projects are not only to result in the digitalisation of material but also in a methodology for similar activities and decisions regarding the remaining parts of the collection.

It goes without saying that the outcome of the projects is useful for others owing to their interdependence and coherence in a larger framework. An example of this is information-retrieval projects.

As I've already mentioned, the (partial) digitalisation of the collection as a matter of fact amounts to the digitalisation of the entire library service or the revision of the library organisation. One cannot simply reduce the latter by describing it as an adjustment or expansion of library and IT-staff adorned with some extra technical training courses.

Especially in the case of projects on electronic publishing, there is usually a lack of organisational tactics and guidance provided in the traditional information chain. This chain is now imploding because separate functions are being intertwined, thus becoming indistinguishable from each other (Ref 4). The author is no longer different from a publisher. Moreover, this is linked with all kinds of copyright problems. It's only fair to admit that in Holland not all the legal bickering over electronic publishing and copyright has been completely solved yet. This is partly due to insufficient jurisprudence.

The lack of useful examples makes it necessary to adjust the library organisation through the complicated method of trial and error. And now for some troubles and traps, and their clearing away something on a particular project that has been quite enlightening in this respect.

6. Practice and problems

One of the most awkward and enduring problems proved to be the access to information sources from every place on the campus. In part this problem was also the reason for the poor usage of electronic information sources.

A short case story:

6.1. A possible approach: which scenario and which considerations?

Our objective is clear: to endeavour towards an accessible electronic library from every working site or desk on the campus inspired by the fierce wishes of students and staff members. But how to reach this goal?

In preparing our pursuit we first had to answer two questions: how is the barometer, the 'market situation' for electronic information sources? Second, how's the terrain, the infrastructure at Utrecht University?

The subscriptions, i.e. the licences we have (about 60 databases in all) can be technically divided into two types: platform-dependent (CD-ROMs for DOS, Windows, Apple Macintosh or Unix) versus platform-independent sources (predominantly online databanks) accessible through World Wide Web browsers or Gopher.

Presently most information sources offered by publishers are platform-dependent ones accessible only via DOS. These sources constitute about 90% of the total amount of sources available at Utrecht University: quite a lot.

However, this needn't be a problem if you realise that 80% of all PCs are DOS-machines. Another fortunate circumstance is the comparative modesty of the required hardware. Simple Intel 386-type machines, purchased some five years ago which nowadays are mainly used as wordprocessors, are still suited for receiving electronic information.

Why not connect those faithful old boxes to the information sources on the network? Why, you're then asking yourselves, wasn't this done before? Why did you have to arrange some special project to this end?

The answer is baffling in its banality: because of conflicting network software. The larger part of our electronic information sources are accessible exclusively via the so-called NFS protocol. Most faculties however use Novell, a protocol that cannot communicate with NFS. This means the use of platform-dependent sources is also limited by the type of network to which the computer is connected, this contrary to platform-independent sources.

How to solve this problem? There are several scenarios.

The first one looks very simple. You do nothing: you flatly refuse to spend a single penny on it. For the time being these electronic services will only be available on specially equipped computers. One just has to wait until all information providers are offering their products in a platform-independent manner, i.e. via the World Wide Web.

There is a slight 'but' to this. By then as a user you'll need quite a powerful computer, the latest high-end machine equipped with a top-notch processor and lots of RAM. If your department cannot afford to buy such a machine, well, bad luck for you. Perhaps in some distant future there'll be higher spending budgets instead of the common annual cuts.

The second scenario is less coarse. You concentrate yourself on configurations meeting the demands of predominant information sources (in our case DOS), taking into account the number of these machines in comparison with the total amount of computers (i.e. 80%). Then you start plodding, trying to rub out the incompatibility problem between the network protocols.

In addition one can try and realise a platform-independent access for the few happy owners of powerful computers, assuming that their number is growing steadily.

We chose the latter option and met with some difficulties: problems we were able to overcome.

6.2. Reality

Erasing the incompatibility between network protocols appeared to be a much tougher ride than we had anticipated. It wasn't a trifle: we couldn't just bridge the gap by means of some kind of ready-made translation program, taking us only a week to build.

To be honest, the project took us nine months. I shan't bore you with our misery — all those wrong and outdated test keys, false network addresses etc. — I'd only like to mention one significant and recurring problem.

We always tend to forget *la condition humaine*, the human factor. It never is a matter of mechanics only. Conflicting views, opposing proposals and alternative solutions — the stuff debates are made of, the gist of many a time-consuming discussion.

IT experts usually don't express much enthusiasm at the thought of a platform-dependent access to an electronic library. 'Why do we have to work on an obsolete solution? We've got the knowledge, gentlemen, the technology to realise a state-of-the-art option.'

Library staff knowing the wishes and needs of most users by contrast take a different stance. They dread experiments with software, fast machines and the like, not interested as they are in promising technical innovations. 'It also works the good old-fashioned way. So why bother?'

The faculties are the third party in the game. In the least irksome case they're compelled to discuss all kinds of tailor-made solutions interminably with their IT-departments.

All in all this doesn't enhance the progress of a project. But eventually, after nine months of patience we reached our goal.

Patience is the key-word for *problems*. The key-word for *solutions*, though, is perseverance. One must never forget the purpose of the project: to comply with the needs of the library's users, i.e. students, teachers and researchers. Very helpful in this respect is the idea of — let me borrow an American expression here — a 'user-centred library'. Most IT experts object to this. Here a different phrase would be better, they say: for instance 'your future research needs are our present concern.'

But we need not polemicise on this. As the above paragraphs prove we're trying hard to realise both objectives.

In this paper I've attempted to outline the framework in which the Electronic Library Utrecht projects are

embedded, plus their governing principles. I've mentioned the example of a complex project to illustrate that this approach offers a guarantee to quite successful results.

We don't reject any utopian visions of information technology — they simply aren't our first preoccupation. We've chosen to give our project's foundations sound and solid ground — just down to earth.

Natalia Grygierczyk
Universiteitsbibliotheek Utrecht
Postbus 16007
3500DA Utrecht
The Netherlands
Tel. +31 30 2536516
Fax. +31 30 2539292
E-mail: N.Grygierczyk@ubu.ruu.nl

References

- [1] Chambers, R. (1844) *Vestiges of the Natural History of Creation*, Wiley and Putnam, New York.
- [2] Negroponce, N. (1995) *Being Digital*, Knopf, New York.
- [3] *Verslag van paneldiscussie over bibliotheekvoorzieningen en -diensten van de Universiteitsbibliotheek Utrecht* (1995), Universiteitsbibliotheek Utrecht.
- [4] Savenije, B. (1995) *De onbegrensde bibliotheek*, Universiteitsbibliotheek Utrecht.



Spanish museum libraries network

Rosario Lopez de Prado

Museo Arqueologico Nacional Biblioteca, Spain

Abstract: Most of the State Spanish museum libraries have valuable and specialised funds which are at the present difficult to manage due to the fact that they are sometimes at different and, frequently, small institutions. The problems seem to be easy to solve through the creation of a library network that will keep these valuable funds from getting lost, and at the same time will help researchers and professionals retrieve information from these collections.

To do this it is necessary to create a 'bus net' with mainframes at Museo del Prado, Museo Arqueologico Nacional and Centro de Arte Reina Sofia. From these sites the programs will be loaded and connection with the libraries will be established according to the ISO/OSI standards. That stage of the project involves the existence of a professional group to coordinate cataloguing work which has previously been done at the libraries concerned.

In the second stage, libraries will be connected with the museums' databases; later on the network will be connected to the online databases at the Ministerio de Educación y Cultura, Consejo Superior de Investigaciones Científicas, the universities, etc. Finally, all the information could be available on the Internet through the international TCP/IP protocols. In this way, documents now at the Spanish museum libraries could be available all over the world.

Keywords: Museum libraries, automated networks, online databases

1. Introduction

All Spanish museums depending on the Central Government have larger or smaller libraries, sometimes considered as individual departments and other times as storehouses for books which serve as a sporadic documentary base for museum work. Many of the large European museums started by being connected to a general library in answer to the universal concept of Culture typical of the 17th Century. Nevertheless, in most cases, both institutions separated and became independent of each other. At the same time authentic museum libraries began to appear, frequently with specialised materials taken from the first general library, made out of collections which serve as a base for and a complementary part of all museum activities, and for the objects stored there.

Little by little, as museums began to define more precisely their functions and limits, libraries depending on museums turned into specialised centres. Unfortunately, in most cases libraries were considered as an annex of little importance to museums, and were denied the prominence that centres for specialised documentation should have had. Traditionally museum libraries were considered just as storehouses for books, of no importance to a museum, and often were neglected from the beginning.

2. Spanish museum libraries.

Nowadays, museum libraries depending on the Central Government contain numerous and different documents — from 80,000 at the Museo Nacional Centro de Arte Reina Sofia to 2000 at the Casa-Museo de El Greco — with monographs and periodicals as a whole, but also containing audiovisual records, microforms, slides and so forth, auction and sales catalogues, exhibition catalogues and old prints. This material is geographically dispersed and access to it is difficult, when not impossible, since many of these libraries do not have specialised personnel in charge and in many cases the material is not even adequately catalogued. This situation becomes complicated because of the different status of each museum. There are autonomous organisations like the Museo del Prado or Centro de Arte Reina Sofia, and others depending on the Subdirección General de Museos Estatales, in the Ministerio de Educación y Cultura. Material held in their libraries can be very different, as can services, norms for access, actual or potential users, budgets for maintenance, personnel assigned, etc.

Nevertheless, the Spanish museum libraries are centres of enormous interest for different reasons. In the first place, documents held in them are of a highly specialised nature, often very difficult to be found elsewhere and generally expensive; these documents are part of the Spanish Bibliographic Heritage but they are not collected in the Collective Catalogue (National Union Catalogue), which means scarce protection and availability. Secondly, if museums are considered as live centres for research then libraries are institutions which we cannot do without,

Online Information 96 Proceedings

Page 417

and are responsible for organising documents in a museum so as to help investigators find the necessary information that they often cannot find elsewhere; libraries now go from collection management to access management. Now, none of the functions can be found in more than a few museum libraries, and even then not in a satisfactory way. Besides, the enormous amount of documents constantly appearing on any subject and constantly growing makes it impossible for any specialised library, regardless of its budget, personnel or space, to acquire, manage and hold all the information about any research field. All the special libraries in the world are unable to do it; all of them are incomplete, and will continue to be so, even though research will demand more information and of higher levels. No single institution could today dream of collecting all that is published all over the world in order to assure its users any document they need. Jacquesson has said that 'the autarchie age in libraries has finished'.

On the other hand, the scarce identity of some of the museum libraries makes it impossible for them to automate efficiently, which entails taking over quality software, buying sufficiently powerful hardware, preparing sites, training personnel, etc. Nevertheless, materials for the museum to which they belong — frequently duplicated — force them repeatedly to buy the same books, sometimes in fact just for occasional use, hence wasting money.

3. Library networks and museum libraries

The only way in which the specialised libraries in the world today can continue to be active and to offer valid information is first, to automate the service they offer and second, to create network libraries with cooperative plans. This network can be configured with different structures (ring, star, bus, etc.) and many strategies may be useful in order to create a cooperative system, but it always must respond to the following requirements:

3.1. High level of funds

The cooperative plan allows each centre to define specifically one fixed area, and only that. The communication channels between centres permit them to complement each other, giving complete cover toward larger fields. Museum libraries are very specialised centres in accordance with their parent institutions (museums covering very specific fields).

3.2. Making resources more profitable

This is one of the most interesting characteristics of cooperative plans. Not only does this avoid duplicated purchases but access to some services — like bibliographic services, subscriptions, training, etc. — can become cheaper and cheaper when they are used (and, of course, supported) by several libraries. The resources of a library can also be used by all of those belonging to the same network, in order to avoid duplication of holdings and functions. Many times, sharing a server and software developed cooperatively was the only way that most libraries could afford the high cost of automation. Museum libraries always have very valuable holdings, but usually they have few resources for management and a very scarce budget for maintenance.

3.3. Time saver

An automated library network means that a collective catalogue exists. Each record is catalogued only once and afterwards it is connected to as many holding records as necessary; the same authority files could be created and used in all libraries working together; circulation modules work in real time with a database which includes many libraries; the personnel in charge of the network may also be specialised in certain tasks or certain kind of documents. There are many specialised documents at museum libraries, and users demand more and more information and more and more precision. This means that the staff should be qualified for specific work.

3.4. Standards for libraries

Libraries working in a cooperative network must use the same software, the same hardware, the same protocols for connections, the same format for bibliographic records and the same authority files; they also could have the same rules for access, circulation and anything else. This is the best way to provide guidelines in order to manage to useful standards for similar libraries. Museum libraries have different rules, different equipment and different budgets; however, they are very similar libraries. A network is a successful tool to get the same library standards and policy.

3.5.

Finally, it will not be necessary to search through a lot of different centres to look for a certain document or to verify that it is not available at all. Museum libraries are geographically dispersed in their local, regional, national or international areas, so the end-users have to visit one institution after another to get all the information about each item that they are looking for.

4. The Spanish museum libraries network

Creating an automated network of Spanish museum libraries means creating an automated system which includes museums depending on the Ministerio de Educación y Cultura that have material of bibliographic interest. This network would have the following characteristics:

4.1. A bus network

This makes connections between centres easier and, at the same time, keeps them independent. A bus network has no switches or repeaters: this means that transmissions from any one station can be received by all the other stations without problem. The network would be supported by various servers — it would need to be seen whether it would be necessary to set up one server in each centre, or if some small museums could share the same server: however, at least one server must be installed at the Museo del Prado, Centro de Arte Reina Sofía and Museo Arqueológico Nacional, which have the biggest libraries. These would have many terminals connected, serving as work places (for consulting or entering information) or as OPAC places (consulting only). Graphical user interfaces must be used for OPAC computers to make it easier for any user to get information. All computers that were part of the network would have access to the OPAC, as well as others installed for external access. There is now a server at the Centro de Arte Reina Sofía, with a Unix operating system and BRS database management software, and there is another server at the Museo del Prado, with a Windows NT operating system. All computers would have access to the PIC (Points of Cultural Information) of the Ministerio de Educación y Cultura, and those installed in libraries would also have CD-ROM readers for consulting external databases and, where appropriate, for incoming records. The software for an information management system would be the same — it is necessary to acquire a program for each server — since this network would work as just one institution with many branch libraries. This would also make it possible to acquire or subscribe to some services once, for one institution, and not once for each library.

4.2. Access to a single database

Only the existing documents in any of the Spanish museum libraries would be loaded in a database, which would function as a Collective Catalogue. This way, from any screen located in any museum it would be possible to know the books in existence in each of the libraries of any Spanish museum, the number of copies available and their situation, condition of loans, etc. Also, the cooperative use of the circulation module should permit reserving, requesting, loan or reprography of documents from any point of a network to another. However, this organisation of a shared cataloguing system and general circulation requires supervision and control programs, plans and development policies, regulation between partners, and very strict rules about cataloguing and classification systems, which provide guidelines for different works. It also needs a control group or person for troubleshooting, keywords, access permission and so forth: the network manager.

4.3. Specialised libraries

Working in a network, it would be possible that each library could specialise in those fields or services which would be related to the museum it depends on, and that cooperative plans could be established for acquisition and for cataloguing. The essential principles are to reduce the amount of original cataloguing and to eliminate duplication of effort. During the last decades, sharing cataloguing systems has been very successful in guaranteeing a good level of cataloguing, and saving expense and effort: this method could obtain a better income from the resources available in each library, and work could be done in a more reasonable way. Finally, by joining shared cataloguing schemes, libraries are able to build their own files of machine-readable records in a standard form which can be transferred to other automated systems or networks, or to successive generations of automated systems. Now, working in this way also lets each library become more and more specialised in different items and tasks.

4.4. Access to other networks

The existence of a serious library network offering valuable information and documents — more than 500,000 books, 10,000 periodicals and an undetermined number of slides, audio records, videos, etc. only at the museum libraries belonging to the Ministerio de Educación y Cultura in Madrid — and some very specialised services mean that other networks can be contacted, and that libraries can participate in international projects or programmes of a wider field. The possibilities from there are innumerable. Most immediate would be access to the Internet, subscriptions to IFLA and other institutions, taking part in European programmes, access to Spanish library networks (like REBECA for public libraries or REBIUN for academic libraries), creating librarian working groups for Spanish museums and — most important — loading the whole database in the Collective Spanish Catalogue of Bibliographic Heritage. In any case, the network always must operate as one institution: the virtual museum library.

5. Development

After a very serious study involving organisations, search of funds, definition of a model structure and choice of standards, requirements and technology, we have decided that in general creating a network of automated museum State libraries will take place in the following way:

5.1. *Acquiring hardware for the servers and computers considered necessary*

At least one terminal for each library, printers and the Ethernet cards necessary for the connections. We have chosen the following equipment:

- **Servers:** Compaq Proliant 1500 5/166 Mod. 4300 Intel Pentium. 512Kb RAM cache, 166 MHz speed, 32 Mb RAM ECC expandable to 256 Mb, 4300 Mb Fast-SCSI-2 hard disk, and network card NetFlex-3P in PCI 32 bits.
- **Microcomputers:** Digital Venturis 5100 Pentium, 16 Mb RAM cache expandable to 128Mb and 850Mb hard disk.
- **Printers:** Canon LBP-1260 Laser, 12 page/minute speed.
- **CD-ROM:** Reader-Player CD-ROM Sony Double speed connection SCSI and compatible with CD-DA, CD-ROM, XA and CD-I formats. SCSI Adaptec 1542 bus ISA card included

5.2. *Acquiring operating system, database manager system and library management software*

Acquiring operating system, database manager system and library management software, which will function according to international standards — IBERMARC format, cataloguing rules, ISO norms for protocols to connect networks, character set, open system interconnections, etc. — and which will be loaded onto the servers. It is essential that software adopt the international norms for documentation and be conveniently tested in other libraries of similar characteristics. We have chosen these software programs:

- **Operating systems:** Compaq SmartStar and Compaq Insight Manager (included in the server); Unix (operating system for server); Windows (operating system for client).
- **Database manager systems and programming language:** INFORMIX Standard Engine V.7 for programming language; BRS for databases manager system.
- **Library manager system:** in order to set up the network we prefer a client/server version of the library manager software. However, this last decision depended on the Subdirección General de Museos Estatales: we only can propose the software we considered the best.

So, we have proposed the following software:

- ABSYS, used in State public libraries in Spain;
- LIBERTAS, used in academic libraries;
- SABINI, used at the Library of the Ateneo (Madrid);
- SIRTEX, used at the Biblioteca Nacional (Madrid).

All are very good systems and have been tested in many libraries in Spain; all of them are able to deal with the following problems:

- (a) transferring bibliographic records from external sources to local files;
- (b) creating original catalogue and editing records;
- (c) maintenance of authority files;
- (d) control acquisition and creating spreadsheet;
- (e) serials control;
- (f) circulation control;
- (g) access to external information systems;
- (h) friendliness of OPAC;
- (i) others: account, statistics, interlending, etc.

5.3. *Establishing the corresponding connections between terminals and servers, always following the ISO/OSI norms for interconnection systems*

However, we need access to the Internet so we also have the TCP/IP protocols; we will require our OSI system to be able to interoperate in a reasonable way with systems based on TCP/IP. We have found a solution in the construction of gateways which can interoperate with either type of system acting as intermediaries between both of them. In order to get this gateway we have asked to join an international programme at the CICYT (Interministerial Commission for Science and Technology). The emphasis of this project is on interoperation between different systems working at museums, special libraries and universities in Spain. In this way, we could

benefit from the technology of this project and also from its results.

For connections and communications in the network we will use the RDSI (Integrated Services Digital Network), which provides digital connections from point to point and supports a wide range of services. The RDSI is able to transmit any kind of information (sound, images, text, etc.), and users can have access to them through many standard interfaces. Using RDSI has many advantages: it is faster, simpler, cheaper and more reliable.

5.4. Catalogue retrospective conversion and database loading existing at the moment

For this, three methods can be used (direct keying; matching and extraction of records from a resource database; or optical scanning and automatic format recognition) in different ways (temporary contracts, agreements with institutions, etc.). For authority records we prefer to use authority records from the Biblioteca Nacional, now available on CD-ROM. But in any case, cooperating procedures should be followed, aimed at avoiding duplicating work to economise time and resources. When retrospective conversion catalogues are made, from contracts for copying of records or through the inscription of bibliographic services, it should be only as one institution with branch libraries and not as various independent centres. In any case, there will be at least one member of staff having specific responsibility for the network and its operations, and a group charged with authority files control, database debugging and providing standards for bibliographic records.

5.5. Training personnel and making the network function

Training personnel would be done gradually as the different stages of the project start to function; personnel involved in the same thing would be trained together, which would reduce costs and let the system develop. In any case, training techniques depend on the resources available but they must include classroom and theory sessions, audiovisual presentations and computer-assisted learning.

5.6. Linked with the general museum system

When museum libraries are automated and established in a network, they can never forget that they are moving to a museum-wide documentation system. In this way, we have planned the following objectives:

- (a) all the museum staff must have access to the library network databases from their own departments;
- (b) most of the museum's objects are studied in many documents from museum libraries. A linked system must let information be moved from bibliographic records to object description records, using the same commands and similar interfaces;
- (c) all libraries working in the network have to use the same thesaurus and the same controlled list of subject descriptions, but the museum must use them too. Terminology is the bridge between museum departments and museum libraries.

5.7. Last step

Connection with PIC, access to the Internet, inclusion in the Collective Catalogue of the Bibliographic Spanish Heritage. Once the system gets started it should be opened up externally to improve its rentability and at the same time to use other available resources online. It is suggested that the first measure be connection to the PICs of the Ministerio de Educación y Cultura; access to the Internet as one institution, with a view to offering information as well as retrieving it; and automated record loading in the Collective Catalog of the Spanish Bibliographic Heritage, with an end to acquiring better protection for the material.

6. Conclusion.

Creating a network for automating the Spanish museum libraries is the only efficient way of using the materials and resources available there and avoiding their loss. Automating all the museum State libraries simultaneously on a low budget, using funds jointly and rationalising their use, obtains better performance from the personnel available and gives access to international information networks. Another alternative means a disproportional investment in software, hardware and personnel or the loss of valuable funds.

The joint automation of the museum libraries and their appearing in a network should be seen together with the general project of automating the museums. But it should be made clear that automating the libraries has problems and peculiarities very different from the museums, because of the kind of funds which they hold and because of the functions between them. Automation of libraries is also a project developed more than 30 years ago, with existing agreements for normalising international rules and guidelines from official organisations with details for the process. There also exists an abundant market for informatic programs, which comply with the required rules tried out in other libraries sufficiently to offer a satisfactory standard of quality.

Starting a network is identified with standards and guidelines upheld by the European Union in matters of information aimed at sharing resources and easing access to the existing documentation of member countries, and to assure information flow between them. Creation of networks in libraries and documentation centres with an active participation in wide area networks is a general trend which follows documentation, and a trend that will take place in the whole world over the next years.

Rosario Lopez De Prado
C/ Serrano, 13
28001 Madrid
Spain
Tel: +34 1 577 79 12 x 226/227
Fax: +34 1 431 27 57

Bibliography

- Barnett, P., S. Clarke and G. Earnest (1992) Museum Library OPAC. Discussion Group in the Twenty Annual Conference ARLIS/NA, *Art Documentation*, **11**(2), 80-81.
- Cacho, C., J. Martí, R. Maicas and M. Armada (1990) *Informatización y Documentación en el Museo Arqueológico Nacional*, ANABAD, Madrid.
- Cano, V. (1995) Information technology and the future of professional library practice, *Library Networking in Europe*, October, 12-14, TFPL, London.
- Clayton, M. (1991) Gestión de automatización de bibliotecas, *Fundación Germán Sánchez Ruipérez*, Madrid.
- Diéguez, F. and X. Agenjo (1988) Pautas para el desarrollo de un programa integrado de gestión de bibliotecas, *Revista Española de Documentación Científica*, **12**(3-4), 339-357.
- Dodds, D. Computer applications in the National Art Library, *Computers and the History of Art*, **3**(2), 15-25.
- Elgaard, B. (1993) Museum librarian: everyone's partner, *Museum International*, **10**(4), 48.
- García Melero, L.A. (1988) Automatizar nuestras bibliotecas, *Boletín de ANABAD*, **30**(4), 393-410.
- Hormulos, P. (1990) Museums to libraries: a family of collecting institutions, *Art Libraries Journal*, **15**(1), 11-13.
- Jacquesson, A. (1995) *L'informatisation des bibliothèques*, Cercle de la Librairie, Paris.
- Kesner, R.M. (1994) The library as information centre: a 'utility' model for information resource management and support, *Library Trends*, **42**(3), 373-394.
- Larsen, J. (1985) *Museum Librarianship*, Library Professional Publications, Boston.
- Marchipont, J.F. (1995) *Les nouveaux reseaux de l'information: enjeux et maîtrise de la société de l'information*, Continent Europe, Paris.
- Marks, K. and S. Nielsen (1991) *Local Area Networks in Libraries*, Meckler, London.
- Martínez, L.J. (1993) Teleinformática y cooperación bibliotecaria, *Revista Española de Documentación Científica*, **16**(4), 341-359.
- Root, N.J. (1980) *Role of the Museum Library*, Smithsonian Institution, Washington DC.
- Seminario sobre Redes de comunicación bibliotecaria* (1994) Toledo, 7 y 8 de junio, 1994, FUNDEBI, Madrid.
- Stork, H.G. (1996) Building blocks for Europe's electronic libraries. The European Commission's Telematics for Libraries initiative: taking stock, in *Proceedings of the Tenth Annual Computers in Libraries International Conference*, London, 20-22 February.



Making the best of new information technologies at Eurostat

Christian Guittet

Commission of the European Communities, Luxembourg

Abstract: Keen to improve its statistical information service constantly by taking full advantage of the most appropriate and modern techniques available, Eurostat, the Statistical Office of the European Communities, is already making use of the new possibilities offered by the multimedia revolution and has started research to allow further progress in this direction.

After first defining interactive multimedia as the achievement of a dynamic symbiosis between several means of expression (i.e. several representations of real or imaginary worlds), this paper draws lessons from Eurostat's operational achievements and explores different areas open to future research: multimedia, new IT, computer graphics and visualisation.

Keywords: statistics, electronic dissemination, SGML, multimedia, information highways, Internet, CD-ROM, computer graphics, image processing

Twenty years of Online Information Conferences ... it's time to look back ... have times changed that much? 1976: computer games were the craze; 1996: computer games are still the craze — but crude ping-pong you could only play on big machines available in public places has now given way to interactive movies you can play at home.

Fifty years of computing ... 1946: ENIAC, the first computer ever, slowly performed its first operation; 1996: mainframe computers which churned out tons of paper covered in text written in capital letters belong firmly in the past, and having on the desk or at home a personal computer which speaks and displays photographs or films now surprises no one, least of all the young, who have long been familiar with such technologies through video games.

The global information society (Ref s 1,2) has become fashionable nowadays and, as we are constantly being told, information highways drive us into the age of multimedia. However, what exactly is that? Unfortunately, although everyone is talking about it, not everyone means the same thing. The 'experts' all give a different definition while the layman speaks of a threefold marriage between television, computer and telephone, with leisure, communication, education, reading, pictures, music etc. all in mind.

Is any attempt to define the concept bound to fail? The very formulation of a definition is open to criticism, as 50% of turnover in five years time will come from products which are yet to be invented while multimedia will disappear, becoming invisible, like today's electric engines.

Aware of these difficulties, I therefore propose for the purposes of this paper to define interactive multimedia as the achievement of a *dynamic symbiosis between several means of expression, i.e. concurrent representations of real or imaginary worlds.*

Symbiosis, in biological terms, indeed refers to a lasting, mutually beneficial association between two living organisms. This is what we are dealing with here. It is not a matter of simply juxtaposing text and numerical data, sound, photographs and film, charts and animation. Several or even all methods of expressing knowledge should be thoroughly combined so as to integrate them in a whole which is more than the simple sum of the parts.

Dynamism is essential as well. The qualification 'interactive', often only implied but in fact inseparable, reflects the dynamic nature of a multimedia product. It is not fixed and unchangeable. Each reader recreates it each day, according to his or her own individual needs and personality, freed by such flexibility from any constraint on how to use it and gain access to knowledge.

The lessons of experience

Quite a number of people think more of multimedia in terms of technology than in terms of applications. This is the wrong approach, however: what really matters is information. The key feature of information technologies is that cutting-edge products of today will be available off-the-shelf as soon as tomorrow, while creating content may prove extremely expensive and time-consuming.

This gives Eurostat — the Statistical Office of the European Communities — and more generally National Statistical Institutes much potential to be major players in the global information society. Indeed, as a result of statistical culture, a huge asset of primary data is available in a structured, digital form. The value of these data

is confirmed by the existence of a large base of customers willing to pay to gain access to it.

To fulfil its mission, which is to provide the European Union with a high-quality statistical information service, Eurostat is therefore willing to take full advantage of the most appropriate and modern techniques available. It has already implemented the new possibilities offered by the multimedia revolution in a whole range of products: CD-ROMs, online services, information kiosks ... ; it has also started research to allow further progress.

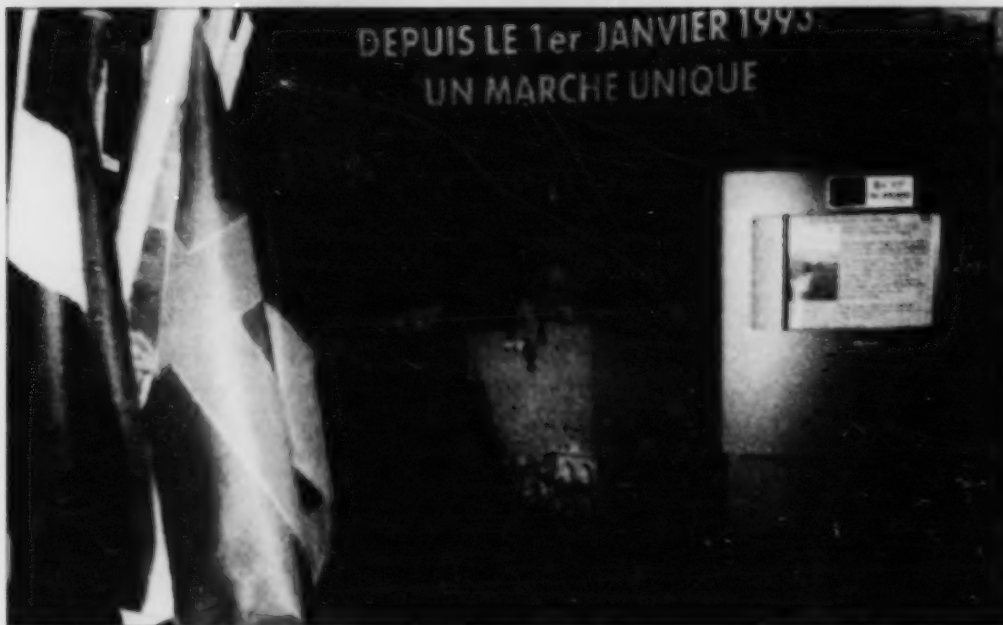


Figure 1: Eurostat Information Kiosk at an exhibition.

For instance, the European Commission's Internet server, Europa, counted nearly 2 million contacts in July 1996, which make it one of the most successful political sites; with 100,000 hits, the Eurostat pages were amongst the most consulted, which clearly confirms the value of the first developments.

Technical solutions adopted

Monolithic developments are definitely out-of-date. To produce this range of applications, the decision was taken to adopt a modular, flexible and open-ended approach, allowing services to be easily upgraded and external partners (National Statistical Institutes, database servers and so on) to add specific information.

More specifically, a combination of different software packages and file formats has been used, as most appropriate:

- Asymetrix Multimedia Toolbook has first been selected as a convenient tool to test different concepts and as an integrator from which other applications can be launched, whenever necessary;
- however, in consideration of the extremely rapid progress of Internet technology and of the flexibility offered by browsers plug-ins, the HTML format now forms the core of these electronic products: Web pages can be distributed not only on the Internet but also on all other appropriate media: diskettes, CD-ROMs, information kiosks;
- Adobe Acrobat has been selected to publish easily and without delay publications for which layout is important or complex, for example press releases (some of which include large tables) or sample pages of publications (even the full publications, for those which are free of charge);
- any presentation package or demo, starting with Java and Macromedia Shockwave applications, can potentially be used similarly to demonstrate electronic products;
- any Internet browser (depending on the user's preference) can be used in conjunction with appropriate plug-ins to view these pages; their design is indeed based on the HTML 3 specifications while extensions such as those proposed by Netscape Navigator or Microsoft Explorer are used whenever appropriate, in a way that does not make pages incompatible with other browsers;
- regarding bitmaps, JPEG files (*.JPG) are preferred to *.GIF files, because they generally allow for higher compression ratios while preserving RGB-true colours; however, exceptions to this rule are accepted for those images which are better compressed in the *.GIF format, as is the case for simple geometric drawings such as some charts, maps or book covers ... or whenever simple animation is needed; the progress of emerging, alternative compression techniques is also carefully monitored;

- regarding sound recordings, the Wave format is being used at present but the intention is to include real-time audio, using technologies such as RealAudio;
- regarding video, tests have been conducted with different alternative formats to take into account the multiplicity of hardware: Microsoft Video for Windows is presently used, in conjunction with the Intel Indeo compression codec; as an alternative, the MPEG standard will be considered when playback facilities become in general use, while emerging technologies (such as Microsoft ActiveMovie or Iterated Systems Coolfusion) seem promising for real-time video.

General guidelines

We have already been able to deduce from these first achievements some general guidelines that we will take into account for any future development:

- *Automate production by structuring information.* This is the only procedure likely to ensure not only the quality of information but also its timeliness and the monitoring over time. Like any other type of publication, multimedia publications have to be updated automatically from reference bases in which the logical structure of the information is identified, using techniques such as SGML. The ability to reuse general texts, illustrations, bibliographies (including the digitised pictures of covers or inside pages) and statistical data is then a real advantage. However, great care must be taken to ensure compatibility with the present situation, however unorganised it is: fortunately, electronic versions of older publications can easily be produced using software tools such as Adobe Acrobat.
- *Ensure independence from the physical medium.* The basic philosophy must be identical whatever the medium (CD-ROM, Internet ... even paper) subject to the limitations imposed by its technical nature. The concept of interactive multimedia transcends the notion of the physical medium with the result that the objective can (indeed must) be to produce similar publications in various forms, with a strong corporate image; one must however acknowledge that the wording itself of articles and texts has to be tailored in places according to the very characteristics of the medium: for instance, hyperlinks clearly have to be considered differently, whether they appear in a printed book or an electronic publication.
- *Be flexible to satisfy users.* Only the best is good enough. But not everybody agrees on what is best, and specific technical limitations make the situation even worse. All the information on offer must therefore be presented systematically in the most appropriate formats, taking account of both production and dissemination aspects; as far as possible, the reader must be presented with alternatives, according to the actual availability of viewers on the various systems used. Using international (*de jure*) or industry (*de facto*) standards should always be considered, if only to benefit from the past experience implied by their making. Off-the-shelf products should also be favoured for the same reasons.
- *Be user-friendly to all users.* And 'all users' really means 'all users': the nerd or computer freak who runs all the latest beta-releases of all Internet browsers in turn and has downloaded all plug-ins; as well as the odd one who hardly knows how to turn his PC on and does not have the slightest idea of which applications have been installed — in fact, he doesn't even care as long as everything works. The system should also not take any unsolicited initiative but be designed to guide the user in need step by step to the 'right' solution.
- *Take care of the user interface.* The quality of the user-interface is of prime importance: it should of course be consistent and designed according to state-of-the-art ideas and best practices, making the service easy to use (ideally, no user-support should be necessary: there is no way to offer a personalised service to thousands of Internet users). But it should also incorporate some nice ideas that will make it attractive; for instance, in Eurostat's catalogue, a click on a cover gives access to the contents, in full in the case of free publications or in part. Examining successful products (Internet sites, CD-ROMs ...), whatever their subject, is always a rewarding experience, although copyright should never be infringed.
- *Avoid sticking to present technology.* The extreme speed of technological progress in this area means that a pragmatic, modular approach has to be adopted to simplify any future development in relation to changes in the environment. Technical choices have to be determined according to considerations such as:
 - the absence of clear, lasting solutions, which can be overcome by the observation of general trends on the market and the gradual emergence of standards;
 - the need to monitor closely the very rapid development of information technology;
 - the heterogeneity of hardware and software in general use, given the need to produce a product capable of very wide dissemination;
 - the cost of dissemination, given the need to pay for licenses for each copy distributed in the case of certain types of software;
 - the need to make rapid progress to avoid diluting the efforts made and the possibility of drawing on past experiences.
- *Put the focus on multimedia resources.* One picture is worth a thousand words. Multimedia resources as

such (music and sound recordings, still and moving pictures ...) are also needed in quantity, although their use must take technical constraints such as the limited bandwidth of networks into account. Several possible sources exist concurrently:

- royalty-free 'libraries' are a good start as they are the easiest to access but they have the disadvantages of being primarily American in origin and therefore culturally biased (for European users at least), frequently used (which may give the impression of *déjà vu*) and not being specific to the message to be put across;
- the use of resources available in the organisation entails selection, followed by processing, but has the advantage of involving other services in the product which will help dissemination; also, prices of appropriate equipment (scanners and video capture cards) and service bureaux (Kodak Photo-CD) are coming steadily down, making such an alternative more affordable;
- more specific illustrations, such as covers and sample pages of publications, interviews, statistical maps and so on add the final, corporate touch to a product.

The key to success: automating the production process

SGML (Ref 3), the ISO standard to mark-up text, has been around for more than 10 years now and I still recall the great expectations that some placed on it when it was introduced: it would finally make publishing fast, cost-effective and easy.

The Office for Official Publications of the European Communities has been a forerunner in Europe, placing SGML at the core of its electronic publication system with FORMEX (Ref 4) as early as 1983 (at that time, SGML was not yet an ISO draft proposal). Eurostat followed the same tracks a little later and developed its STRINGS system. This SGML-like approach to automating the production of statistical documents is presently being upgraded to offer more advanced facilities and to take multimedia publications into account; the new release will be fully SGML-compliant.

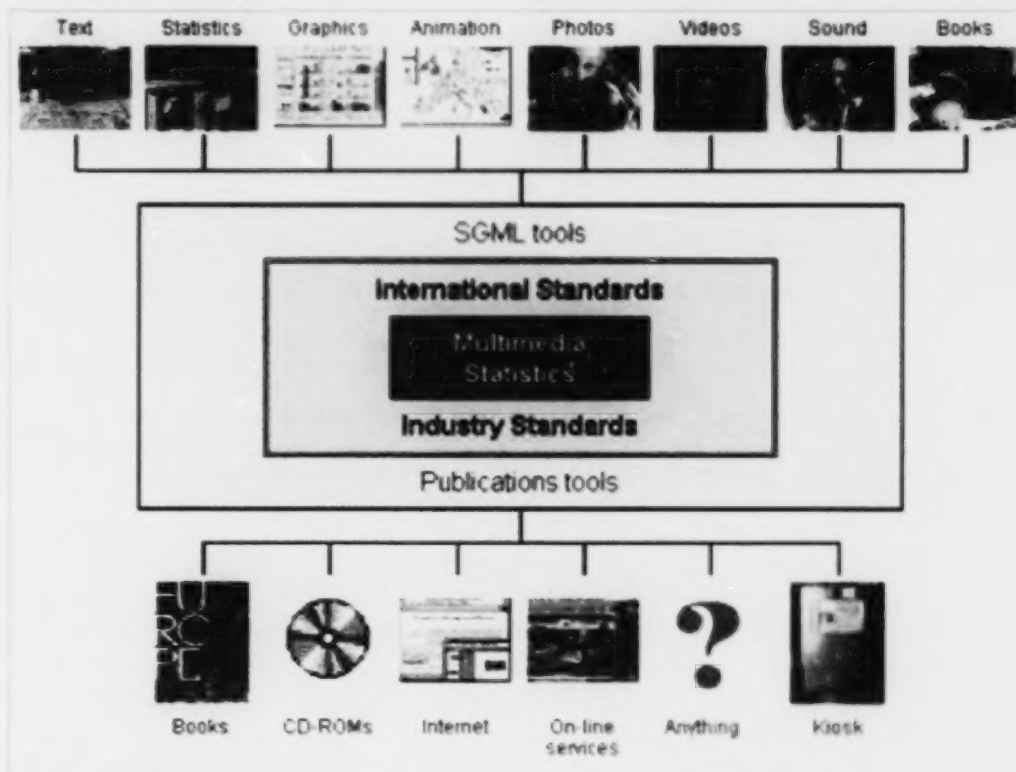


Figure 2: Multimedia statistics.

All over the world, interest in SGML is increasing steadily but slowly, so that the rumour may be true that this acronym really stood for *Sounds Good, Maybe Later*. Indeed, I still know quite a few people (whose name I won't disclose here) who are not only unconvinced but even refuse to hear this four-letter word. They are afraid and still believe SGML is complicated, with all these markers and declarations — it is pretty unfortunate that possibilities to hide all the codes (using advanced features such as *OMITTAG* or *DATATAG*) have not been better advertised and that real SGML editors are not more widely available.

But things may change rapidly now, as these people have discovered a new magic keyword: not *abracadabra* — electronic publishing is not a fairy tale — but *Internet*: they are quite sure that the Web will make everything simple, as if by magic, and are volunteering to become the gurus of HTML ... while SGML-like editors are finally becoming popular.

However, a number of issues are still open and they should not concern Eurostat only.

Is SGML still needed?

Starting to use SGML generally implies a lot of effort to modify the production process radically. If your intention is only to produce one-off paper publications, such an investment is probably not worth it. If you want to enhance these publications by also distributing their electronic copy, using SGML is probably still not worth it; it might even be a disadvantage, as SGML does not record the layout of a document: although the electronic version will be identical in content it will not resemble the paper original.

In such a situation, a better and easier solution is certainly provided by tools like Adobe Acrobat, which in an hour or so will turn any publication, produced on any system, into its electronic 'photocopy', while preserving access to the textual information itself (copy and paste, search and so on) and even allowing for enhancements (hyperlinks ...).

However, if you want to distribute a lot of information in different forms, targeting different users and making the best of all the facilities provided by the different technologies, you should probably start by structuring your information and storing it in a database; you would then be able easily to generate files marked-up as appropriate and use stylesheets actually to produce a whole range of publications in an automated way. In such a case, SGML is certainly a worthwhile approach, well worth being considered.

Is HTML a valid SGML application?

Unfortunately not! For the time being, at least! The key characteristic of SGML, which gives it value, is that layout information is never embedded in a document: what really matters is its logical structure only, while formatting is defined in associated stylesheets, one for each medium.

This is unfortunately not the case with HTML: markers such as *<ITALIC>* or *<BOLD>* are obviously inappropriate and should be replaced with ** (emphasis, typically rendered as italic — but not necessarily) or ** (strong emphasis, possibly rendered as bold) which are also recognised by HTML browsers; but tags such as *<CENTER>* or attributes such as *<P ALIGN = Justify>* have no acceptable equivalent; also, it is quite unfortunate that using attributes such as ** should be recommended for practical reasons when writing pages for the Internet. Introducing stylesheets is urgently needed to improve the situation (and the news is good).

How should you define your documents?

The DTD (Document Type Definition) is a key feature of SGML, used to describe the logical structure of the information. But this structure can only be defined in relation to the needs that must be met. For instance, suppose you want to publish statistics on unemployment. If all you have in mind is a table showing short-term and long-term unemployment, by sex and age-class, with one row per country or world region, you will probably consider a declaration defining a table (*<TABLE>*) as the combination of a caption (*<CAPTION>*), rows (*<TR>*) and cells containing a header (*<TH>*) or data (*<TD>*); this is the HTML approach and it will probably work fine. As long as you do not need to know what information is precisely contained in the cells, that is.

Indeed, in doing so you loose all semantic information related to the significance of the figures and any hope of processing this data: no calculation is possible, nor any charting of data. Your customers will not be able to take advantage easily of the facilities now offered by interactive multimedia.

The way ahead

In order to disseminate statistical data to as large and diversified a public as possible, it is important to make the most of the possibilities new information technologies offer to ensure easier, faster access to more relevant information, according to the knowledge, expectations and needs of each individual.

Many people are in fact likely to use or be interested in statistical data. However many of them, particularly the non-specialists, may be put off as much by the difficulties of access to relevant, up-to-date information as by the daunting appearance of tables of printed figures, just as the opposite is true, as the success of Eurostat's flagship publications shows (*Facts through Figures*, *Europe in Figures*, *Portrait of the Regions*, etc.). It is however possible to make information a lot easier to find and retrieve by allowing the user to explore and understand what is available and helping him to formulate his requests, and a lot easier to comprehend by presenting results adequately.

Also, it is generally recognised that the key to progress is to establish links between (apparently) unrelated application domains and technologies — or, to be more provocative, between virtual user-needs and virtual technologies. The development of technology demonstrators can help potential users to identify the fulfilment of their needs as being within reach; the formulation of new needs as a reaction to these prototypes can in return stimulate future progress.

Technology, an answer to user problems

Technology has no value in itself: it is only a tool to respond better to user needs. It is therefore essential to identify specific needs and to relate them to possible solutions; here are just a few examples:

- marketing teams are typically anxious to know their customers better; posting a registration form on the Internet is pretty straight-forward but you then have to rely on customers to fill it in; some incentive might be useful but candid users may simply be unable to answer technical questions — e.g. questions about the specifications of the system used (which processor? how much RAM? which screen? which sound card? what about the CD-ROM? the Internet connection? the browser and its plug-ins? ...); could an ActiveX or Java applet find that out automatically and return the information together with the registration form (subject to user agreement, of course)?
- edutainment is a big success nowadays and its value in increasing general culture is generally recognised; attractive statistical applications can be imagined and programmed for the Internet, using multimedia authoring tools like Shockwave; what about:
 - a quiz where the player would be presented with a blind chart (for instance, the GNP per capita) and be asked to add names of countries as appropriate?
 - a quiz where the player would be asked to fill 'tubes' with a coloured 'liquid' to represent what he believes to be the correct data for different countries?
 - a game where the player could compare himself to the 'average' European or National (for selected data such as: number of children, equipment, income, taxes ...)?
- bringing many people together to attend a lecture by a key personality or to participate in a conference may be a worthwhile exercise, but both time-consuming and pretty expensive; could information highways provide a cost-effective solution, giving large numbers of people an opportunity to exchange experience and improve their skills?

Such practical applications are obviously considered by Eurostat in relation to priority themes, such as the European Monetary Union (Ref 5), education (1996 has been declared the European Year of Lifelong Learning (Ref 6)) and Euro-Mediterranean cooperation (Ref 7).

The Fourth Research and Development Framework Programme (Ref 8) has given Eurostat the opportunity to launch several research activities in this direction. Their general purpose is to investigate leading-edge technologies and to open ways for their future application to the field of statistical information, while staying close to the real needs of statisticians and statistics users. Prospective studies and practical considerations are being combined while carrying out this research, in order to:

- identify emerging technologies of potential interest to Eurostat and the European Statistical System precisely;
- select statistical fields likely to benefit from these new methods;
- define the methodology with which to assess the value of these new possibilities and their acceptance by users;
- create prototypes applying these new possibilities to selected statistical areas in order to test them on a representative sample of users;
- draw up concrete proposals on the exploitation of the results and define new directions for subsequent action.

Contribution to standardisation activities, as appropriate, and publication of results are considered to be important objectives for these projects so that European companies really can benefit from these research activities.

Multimedia and new information technologies (Refs 9,10)

The merging of computing and communication is a key trend in information technology and has great potential. The general objective of Eurostat is to review leading-edge technologies and to evaluate how, in the medium term, they could help further improve the statistical information service provided to the European Union.

For instance, users might well soon be able to access information whenever they need it without any effort: properly equipped, they would invoke and instruct intelligent agents to monitor their own choice of statistics and make contact when specified conditions are met. They would then receive a limited volume of fresh information, combine it with large time series available locally and process the whole set with local, advanced multimedia systems.



Figure 3: Example of slow frame rate video transmitted over the GSM digital cellular telephone network (Ref 9).



Figure 4: Example of animated map: here, the inflation rate — successive periods can be seen in turn while captions and actual data are shown interactively by pointing and clicking (Ref 9).

The following possibilities are being assessed from the point of view of statistical data:

- performance of personal data assistants (PDAs) and communicators (with cellular modems) to access statistical information;
- usability of intelligent agents (knowbots) to retrieve relevant statistical information automatically;
- impact of higher bandwidth networks (e.g. ATM) on the dissemination of statistics-related information;
- possible synergy between networks and local storage media (hybrid CD-ROMs);
- feasibility of decentralized multimedia interfaces for interrogation and visualisation;
- applicability of remote programming and real-time interactivity with multimedia applications on the Web (e.g. ActiveX, Java or Shockwave);
- applicability of virtual reality (VRML) to improve access to statistical information;
- pilot applications of virtual conferences on the Internet;
- pilot applications of real-time video-conferencing on the Internet;
- pilot applications of real-time audio or video broadcasting on the Internet.

Computer graphics and visualisation (Refs 11,12,13)

Computer graphics are already extensively used in other sciences (mathematics, physics ...) to represent realities that cannot directly be observed. Statisticians are however confronted with similar problems: could new computer graphics and image-synthesis technologies be applied to this field as well? The challenge here is to represent simultaneously as much data as possible, as clearly as possible, in order to increase information and to make it more understandable.

A flat image has two dimensions; the third dimension is traditionally represented by projection; computers can add a fourth dimension (time) by means of movement. However, is it possible — and if so, how — to go further in order to represent a large mass of correlated information better? Can new methods of data representation, whether static (typically, printed on paper) or dynamic and interactive (typically, displayed on a computer screen) enhance the user's understanding of statistics by:

- really introducing the third dimension, by the synthesis of stereoscopic images or holograms?
- really introducing the fourth dimension (time) by animation?
- using the three components of colour (RGB or HLS) to visualise three extra dimensions?
- transforming the original data, using advanced mathematics, before visualising it?
- using interactivity and virtual reality techniques, to go one step further?

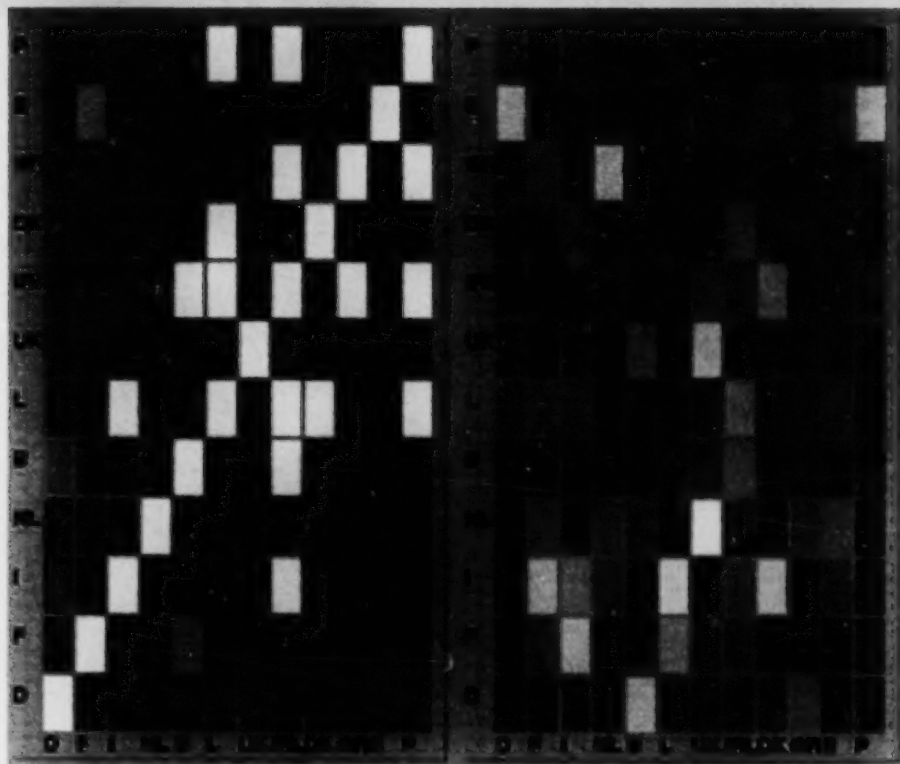


Figure 5: Colour Table Visualizer (Ref 12), applied to international traffic (1993) by road, rail and inland waterways.

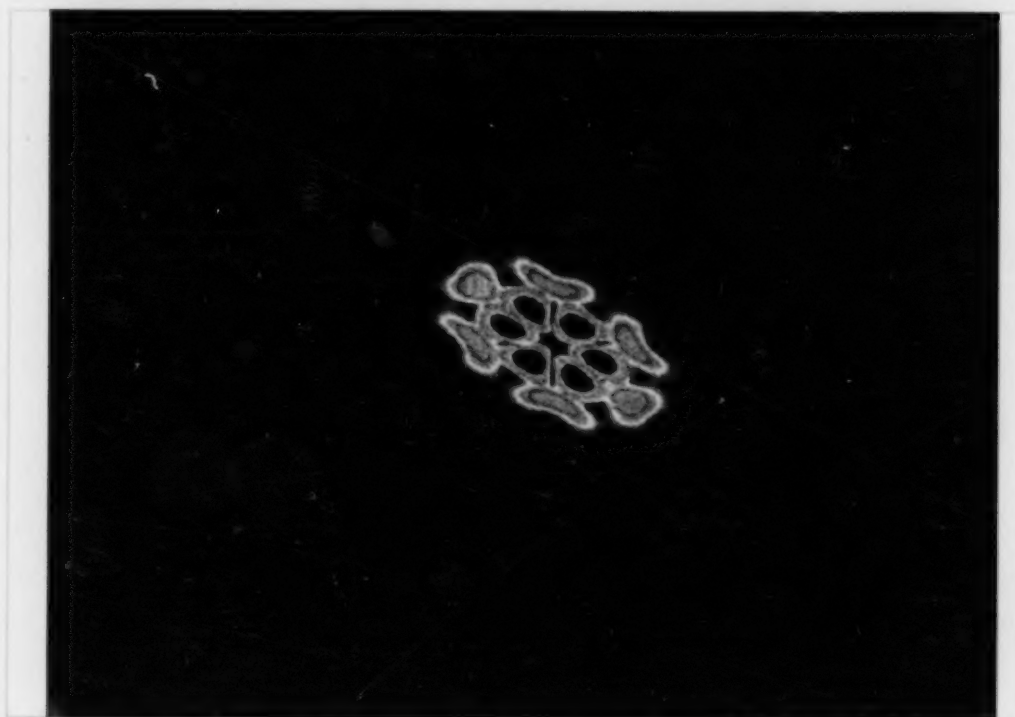


Figure 6: Bi-spectrum of the quantity of sports goods manufactured in Germany over the period January 1978–November 1995 (Ref 14).

The following possibilities are being assessed from the point of view of statistical data:

- possibilities to represent graphically as much data as possible, simultaneously and as clearly as possible, in order to make information more understandable to the user;
- possibilities to represent graphically as much data as possible, simultaneously and as clearly as possible, in order to increase information by letting unsuspected correlation appear;
- suitability and implications of applying mathematical transforms, such as parallel coordinates, to data prior to their graphical visualisation;
- applicability of advanced techniques, such as stereoscopic images, holograms, interactivity and virtual reality ...

Image and signal processing (Refs 13,14)

Ideally, multi-dimensional data should be presented so that all of the important information is accessible quickly. Over the last decade there have been significant developments in the fields of signal processing, image processing and video compression. This progress has resulted in a number of powerful techniques for the manipulation of multi-dimensional data. In addition, a number of powerful nonlinear data compression techniques have recently been developed for such data in the field of neural computing. By retaining only the most important aspects of the original data (the 'real' information) they might enhance intelligibility of complex data sets. However, to be efficient they generally exploit some *a priori* knowledge of the data structure.

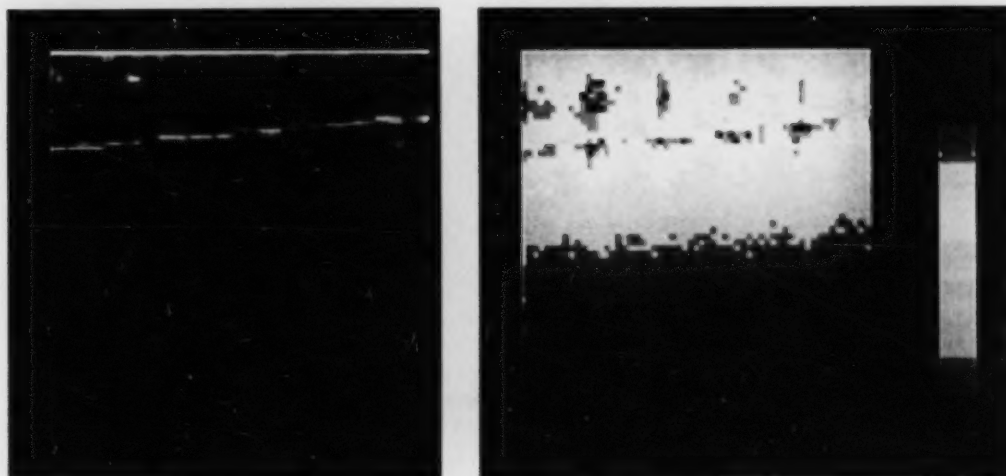


Figure 7: Application of horizontal edge detector (left) and pseudo-colouring (right) to heart attack mortality data (Ref 14).

The overall objective of this work is to identify new methods of representing statistical data which are based on signal processing, image processing and data compression techniques, and to evaluate how effective these new techniques are when compared to established methodologies. It will in particular:

- consider evolving and established signal processing, image processing and data compression techniques which will ensure that a broad range of approaches are addressed;
- compare the usefulness of the new techniques with that of established methods and thus provide a clear and representative evaluation of the strengths and weaknesses of the new techniques.

These questions raise the issue of precisely defining what is information, although it closely depends on individual users' expectations. For instance, one statistician will be interested in time series to analyse their trend with a view to forecasting the future, while his colleague next door will on the contrary focus on trying to understand and explain their spiky components; the noise then becomes the relevant information.

Results of this research are not yet available at the time of writing this article (September 1996); they will in principle be published in a summary report detailing the methods tested, the results obtained and guidelines for the future. This publication is likely to be made available on the Internet, together with some prototypes, so that feedback can be obtained from potential users. Results will also be presented at conferences like this one, when available.

Crystal-gazing

Crystal-gazing is a dangerous activity, especially in the field of new information technologies where progress is so fast. Reviewing predictions that concluded presentations at the 1st Online Information Conference, some 20 years ago, would undoubtedly be great fun.

I shall run the risk, nevertheless. What will be the trend for the next few years? A brave, new world is coming, where all problems will have been solved:

- images, whether still or animated, will become increasingly important: bandwidth limitations will indeed gradually be overcome;
- users will interact physically with statistical data: more and more powerful computers, associated to more and more sophisticated visual display units, will give access to virtual reality techniques;
- relevant information will be delivered automatically to users: you will no longer have to spend hours searching for data you desperately need; robots will learn what you are looking for and find it for you;
- you will always have access to the right information: your communicating personal data assistant, in your pocket, will take care of that.

But please allow me not to rate the probability that all this will occur ...

Christian Guittet
Commission of the European Communities
Eurostat A/1 - Computer processing
L-2920 Luxembourg
Tel: +352 4301 32375
Fax: +352 4301 32594
<http://europa.eu.int/eurostat.html>
E-mail: christian.guittet@eurostat.cec.be

References

- [1] *Growth, competitiveness and employment: The challenges and ways forward into the 21st century*. White paper (Secretariat-General, European Commission, 1994 - Catalogue number: CM-82-94-529-EN-C).
- [2] *Europe and the Global Information Society: Recommendations to the European Council* (Bangemann Report).
- [3] ISO 8879-1985: Information processing systems. Text and office systems. Standard Generalised Mark-up Language (SGML).
- [4] Guittet, Christian (Ed.) (1985) *FORMEX — Formalised Exchange of Electronic Publications*, Office for Official Publications of the European Communities. ISBN: 92-825-5399-X
- [5] Treaty on the European Union.
- [6] Decision No. 2493/95/EC of the European Parliament and of the Council of 23 October 1995 (Official Journal of the European Communities No L256/10, 23/10/95).
- [7] Barcelona Conference on the Euro-Mediterranean partnership (27-28 November 1995): involving the 15 Members of the Union, 11 Mediterranean nations (Algeria, Morocco, Tunisia, Egypt, Israel, Jordan, Lebanon, Syria, Turkey, Cyprus and Malta) and the Palestinian Authority.
- [8] Fourth Framework Programme of European Community activities in the field of research and technological development and demonstration, 1994-1998: Council decision 1110/94/EC of 26 April 1994 (Official journal L 126, 18 May 1994).
Adjustment of the budget following the accession of the three new Member States (Austria, Sweden and Finland) on 1 January 1995: Revision to the Council Decision 616/96/EC of 1996-03-25 March 1996 (O.J. L 86, 04 April 1996).
- [9] Statistics, Multimedia and New Information Technologies (Project SUP.COM 95, lot 21; ERA Technology, with City University): <http://europa.eu.int/en/comm/eurostat/research/supcom95.21/intro.htm>.
- [10] Multimedia and new IT, including standardisation activities (Project SUP.COM 96, lot 29; contractor being selected): <http://europa.eu.int/en/comm/eurostat/research/supcom96.htm>.
- [11] High Resolution PC Graphics (Project SUP.COM 95, lot 10; ATKOSoft S.A.): <http://europa.eu.int/en/comm/eurostat/research/supcom95.htm>.
- [12] Image Synthesis (Project SUP.COM 95, lot 27; ATKOSoft S.A.): <http://europa.eu.int/en/comm/eurostat/research/supcom95.27/intro.htm>.
- [13] Computer Graphics and Visualisation; Application of Parallel Co-ordinates to Statistical Data (Project SUP.COM 96, lot 30; contractor being selected): <http://europa.eu.int/en/comm/eurostat/research/supcom96.htm>.
- [14] Evaluation of the Use of Signal Processing, Image Processing and Data Compression for the Analysis of Statistical Data (Project SUP.COM 95, lot 28; ERA Technology, with University of Surrey): <http://europa.eu.int/en/comm/eurostat/research/supcom95.28/intro.htm>.

Online Information 96 Proceedings

Page 432



View-based searching systems — progress towards effective disintermediation

A. Steven Pollitt, Martin P. Smith, Mark Treglown and Patrick Braekevelt

The Centre for Database Access Research, University of Huddersfield, UK

Abstract: *This paper presents the background and then reports progress made in the development of two view-based searching systems — HIBROWSE for EMBASE, searching Europe's most important biomedical bibliographic database, and HIBROWSE for EPOQUE, improving access to the European Parliament's Online Query System. The HIBROWSE approach to searching promises to provide significantly more effective information retrieval for end-users than is possible through simple keyword, command line, forms-based or hypertext linking interaction. View-based searching makes extensive use of knowledge structures in the form of thesauri and classification schemes to provide linked browsable subject views onto databases. The result is a rich interface where queries can be satisfied by selective progressive refinement and expansion of mutually dependent views. The effect for the user is to significantly increase searching power without a commensurate increase in user effort, thereby reducing the reliance on intermediaries for sophisticated searching.*

Keywords: Disintermediation, view-based searching, end-users, user-interfaces, thesauri, classification, EMBASE, EMTREE, EPOQUE, EUROVOC, HIBROWSE

1. Introduction

'Disintermediation' first appears in the INSPEC database in the translated title of a conference paper presented in Milan in 1983 (Ref 6) — 'Microelectronics and teleinformatics as economic and social disintermediation tools. (Elimination of intermediaries between producers, distributors and consumers).' Subsequently it appears in only another eight records on the INSPEC database, yet the issue threatens to cause the most dramatic social changes that will affect the jobs people do since the industrial revolution. The word count for 'disintermediation' using the Alta Vista Web search engine was 665 on the 21 June 1996. The first ten sites mainly concern banking and electronic commerce, yet the third site in the list is 'Online Information 96: Call for Papers'.

Research and development at the University of Huddersfield has been directed towards enabling end-users of retrieval systems to make effective use of databases without the assistance of search intermediaries. All these efforts have made use of database specific thesauri where the user is able to apply the power of recognition and selection as the key characteristic of query specification. Expert system techniques (Ref 8) such as CANSEARCH (Refs 8,10) promised much in the early eighties but proved expensive to build and inferior in performance to the human intermediary. Automatic search statement generation from thesaurus-based menus (Refs 11,12) provided greater generality, more economic development and more powerful searching. This approach required the user to examine search statements describing sets of documents from combining search term selections and did not take full advantage of the user interface to assist the user in their search process. Amending the search statement was not straightforward and the general nature of the interaction had more affinity with batch rather than interactive processing.

This paper describes the most recent developments which evolved from the menu-based approaches in what has been termed 'view-based searching'. This approach was likened to the use of optical coincidence cards (Ref 13) because of the seemingly parallel nature of the processing, yet it offers considerably greater flexibility in the content and filtering options than could have been conceived of when these cards were employed over 30 years ago. The user is now provided with much more opportunity to examine the database and to apply powerful searching as we can simultaneously present several views and employ them to examine the contents of the database by refinement and expansion of different searching elements. Recently this work has been directed at the subjects of medicine and politics and law, through the development of HIBROWSE systems for EMBASE, the major European biomedical bibliographic database, and EPOQUE, the European Parliament's Online Query System.

It is not yet clear to what extent this approach liberates the user from knowledge of the workings of the retrieval system. The paradigm shift to searching through reciprocally refining views has moved the interaction between user and machine further towards the subject matter and away from the operation of the system. This causes the system designer difficulties in modelling tasks and human-computer dialogues (Ref 17).

Online Information 96 Proceedings

Page 433

2. The need for knowledge and skills in information searching

Education and training that enables an individual to play a specialist part in society has been a feature of our existence for thousands of years; no doubt as hunter gatherers we had those who were better at hunting than gathering and vice versa. The division of labour has sought to provide economic solutions to the production of goods and the provision of services. The case for intermediaries to undertake searches for information on behalf of researchers was made by Ranganathan in 1964 (Ref 15):

'The present rate of increase in the number of documents makes it wasteful to leave literature search in the hands of the research workers themselves. There is need to conserve the research potential by a division of labour. In this division of labour, literature search falls to the share of the library profession.'

This reflected the level of knowledge and skills required and the 'unproductive' research time that would be wasted in the days before online searching. Since the onset of computerised searching there have been efforts directed at making it possible for end-users to search for themselves, through approaches which reduce the levels of knowledge and skills required. These originally concentrated on the mundane but fundamental problems of connecting to a searching system but have subsequently moved onto the removal of the need to learn a command language, and now to tools which assist the user in expressing the concepts of queries.

A paper at the Online Information event in 1985 (Ref 9) advocated the pursuit of the 'What' rather than the 'How' in designing interfaces for the end-user in information retrieval, together with a separation of the logical from the physical, if we were to make the most of available technology. The ease of use was likened to travelling from A to B in a taxi, being concerned only with the destination rather than the route.

This question was revisited by David Nicholas in his paper at Online Information 95 (Ref 5). His study, comparing the searching of FT PROFILE by journalists and intermediaries at the Guardian newspaper, discovered that journalists use minimal searching features (only 52% of the commands available) yet they remained generally satisfied with their searching 'thanks largely to the fact they are better at sifting through data in pursuit of relevance.' The command-based interaction here can be said still to depend on the 'How' rather than the 'What' for effective interaction. There was a suggestion that there was also a division of searching where the more complex searching was delegated to intermediaries.

The success of the World Wide Web as a medium for end-user access is in some respects due to the minimal level of knowledge, skills and training required. The hypertext mode of interaction exploits our ability to recognise and all the user has to do is select. This interaction has completely done away with the 'How'. Unfortunately the following of hypertext links as currently provided is unlikely to bring effective information retrieval. More traditional keyword searching through the significant search engine developments, such as Alta Vista, has introduced different levels of complexity for searching that require additional skills and knowledge for effective use. Classified indexes are also being used to help the end-user. The introduction of organisation to any information collection, both gathering associated information together and facilitating browsing, has been an investment intended to save time and improve search performance on the part of the searcher. This is a natural response to the disorganised resource we would otherwise be faced with.

There are still those who would have us believe that natural language queries and ranked output offer the panacea for end-user searching. The size of result sets, and the potential for organising these according to different views, suggests a continuing role for classification as de Grolier stated some thirty years ago (Ref 3):

'We feared some years ago that classification was becoming useless, that the treatment of natural language texts by machine ... would replace classification. Classification and the classificationists would become something like the dinosaurs, killed by the progress of evolution. This has proved to be a complete fallacy. When you examine the new literature you find that more and more classification ... is considered as something quite essential in information retrieval ... It is quite evident that hierarchies, generally speaking, are something which cannot be avoided in an information retrieval system which is to be useful for the reader.'

The increased use of knowledge structures will be seen to become more important as they help in the building of user interfaces which minimise the skill but maximise the power available to the user to search databases.

3. A faceted organisation of knowledge

In the subject area of medicine the application of view-based searching relies on the professional knowledge of the end-user to progress a search. The use of facets in thesauri provides not only a level of organisation but the key to a form of processing which makes the task of information retrieval one of refinement or expansion of component facets, in the discovery of potentially relevant records.

The suggestion that a faceted classification scheme should be used as the basis of all methods of information retrieval was made by the Classification Research Group in 1955 (Ref 1). Forty years later the potential for faceted thesauri in the new century is described by Ron Davies (Ref 2) including the application for searching free text:

'It also suggests new ways to search full text databases, where only the words actually used in a document by its

author are searchable. If a thesaurus with a wide variety of entry terms is linked to a document database, it can provide a way of mapping the wide range of terms used in natural language to a single concept in the controlled vocabulary. A user entering a word in a search expression can have the search expanded to include all of the other possible words or expressions that are related in meaning automatically, thereby providing search results with greater recall.'

Assessing the usability of view-based searching system prototypes (Ref 16) forced a redesign of the user interface to utilise the faceted nature of both the thesauri and the user queries. A fixed set of views where the thesaurus provided the central view (Ref 14) placed too high a cognitive load on the user and did not lend itself to the flexible interaction sought by the designers. Much of the system state required by the user to make the system predictable (Ref 4) was hidden, and although the initial user interface provided means whereby the user could determine the set of values and terms stored in the system to be used in progressing the query, this set is typically too large to be held in the user's memory. The redesign locates the information needed to understand the behaviour of the system in the display (Ref 7).

4. HIBROWSE for EMBASE

The first demonstrable view-based searching systems provide enhanced access to the six million-plus record EMBASE database of biomedical bibliographic records. Assessing the usability of view-based techniques for EMBASE is a project, principally funded by the British Library Research and Innovation Centre, which involves collaboration between the database publisher, hardware and software providers, a database host organisation and a university centre for research and development. Multiple views using the EMTREE thesaurus of 38,000 terms can be accessed and combined with other views, such as year of publication, as required by the query.

In the following example interaction a user is searching for information on the therapy of Alzheimer Disease using the prototype HIBROWSE for EMBASE on the ADABAS system, which at the time contained close to 300,000 records.

4.1. Finding a view

The user selects the diseases facet from the drop down list of views available (Figure 1) and has the option either to browse from the top-level thesaurus terms in this facet or to find a view by keying the first characters from the term, or a synonym, and selecting a term from a permuted list. In this example the user types 'alz' and selects 'alzheimer disease' (Figure 2).

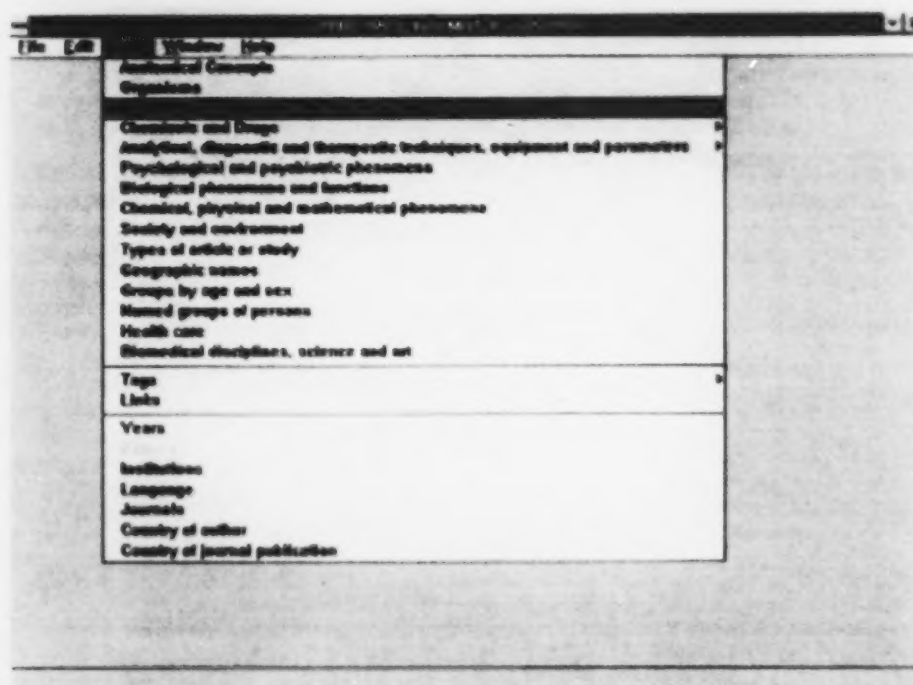


Figure 1: Selecting a facet.

4.2. Refining a view by adding a second view

The user wants to reduce the number of records and decides to refine the view in Figure 3 through the selection of a second view concerning 'therapy' from the drop-down list of facets. A choice is made to browse from the top level thesaurus terms (Figure 4).



Figure 4: View refinement through the addition of a second view on therapy.

The number of records concerning 'alzheimer disease' has now reduced from 1297 to 262 as the second view on 'therapy' has acted as a filter onto the 'degenerative disease' view. Similarly the 'therapy' view itself has been refined to present only those documents which concern 'degenerative disease'. The original unfiltered view is presented in Figure 5.



Figure 5: The unfiltered view of therapy.

4.3. Refining a view by selection from an existing view

Selecting 'drug therapy' on the new view, and electing to refine the views further, reduces the number of records being considered down from 262 to 190, although the numbers of records in the drug therapy view refer to the full set of 817 degenerative disease records (Figure 6).



Figure 6: Refinement through selection within an existing view.

The final refinement comes about when the user refines the degenerative disease view to contain only 'alzheimer disease' (Figure 7). All the thesaurus terms from the drug therapy view which fail to reference any of the documents in the current set are removed from the refined therapy view and the user chooses to see the three titles of those documents concerning 'hormonal therapy'.

BEST COPY AVAILABLE



Figure 7: Final views of alzheimer disease and drug therapy.

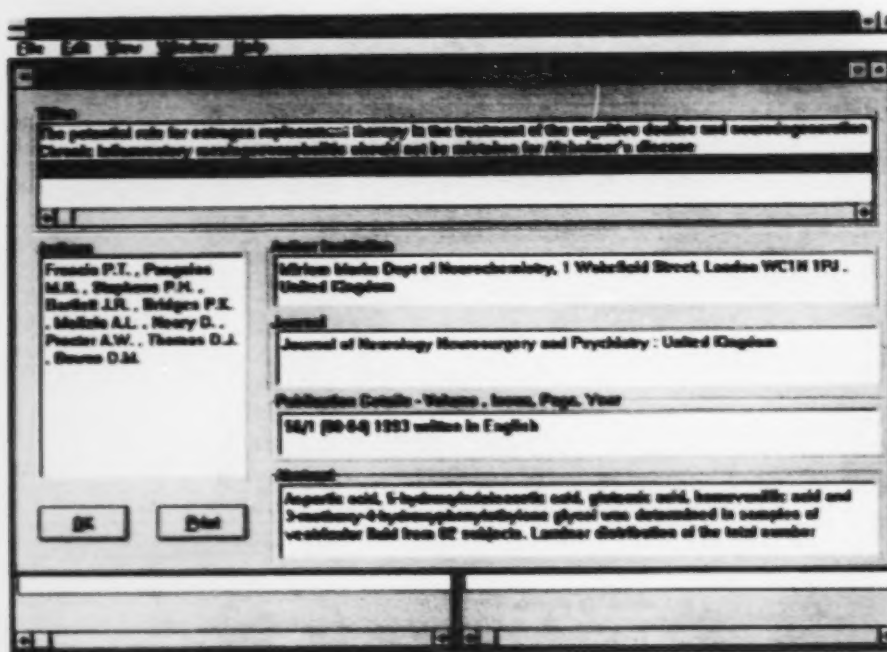


Figure 8: Records concerning Alzheimer disease and hormonal therapy.

4.4. Expanding a view

Should users find the number of records drops to near zero they are able to expand either or both of the views, relaxing the filtering constraints and returning additional terms and associated records to the views.

4.5. Additional views

Further developments will provide views onto the database via authors, author institutions, journals, language of article, country of publication and other attributes of the bibliographic record. The provision of such views will considerably enhance the functionality available to the user without a commensurate increase in complexity. Such features facilitate database analysis as well as high performance searching. Figure 9 shows the additional views provided from the country of journal and country of author.



Figure 9: Adding the views Country of Journal and Country of Author.

5. HIBROWSE for EPOQUE

HIBROWSE for EPOQUE, the document database of the European Parliament, takes advantage of the multi-lingual EUROVOC thesaurus and demonstrates the global potential for improving access to databases where English is not the first language of the user. The following example illustrates a second prototype development which has explored the use of in-line expansion of thesaurus views and the use of additional icons to provide a more informative view. This client software development will also provide access to the full EMBASE database on the STN-International network.

The following example explores the subjects of 'Employment and work' and 'Business and competition'

5.1. Selecting a view and browsing through in-line expansion of concepts

The view on 'Employment and work' is updated online to the European Parliament where documents are added to the EPOQUE database on a daily basis (Figures 10, 11).

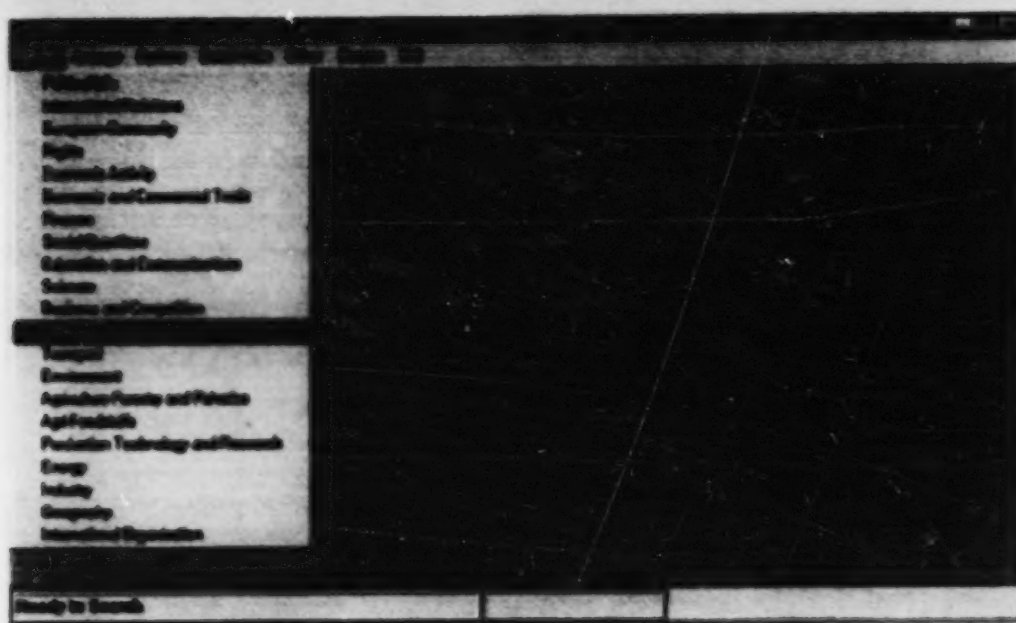


Figure 10: Selecting a view on HIBROWSE for EPOQUE.

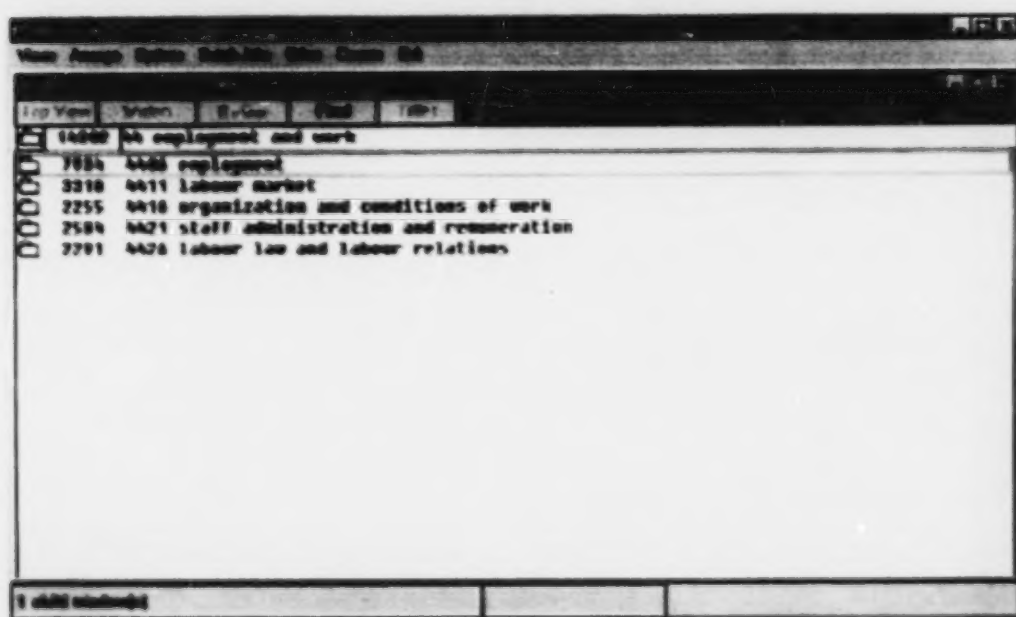


Figure 11: The top level view on 'Employment and work'.

The user can browse these documents by opening folders. Figure 12 shows how the view is expanded *in situ* revealing a breakdown of the 7934 documents into more specific headings.

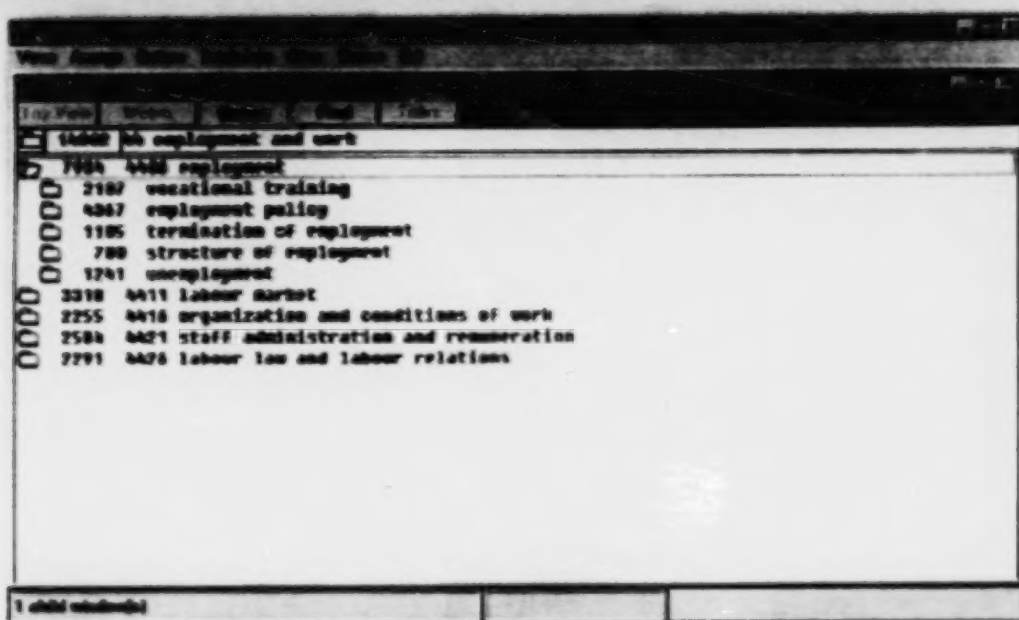


Figure 12: Expanding a view by opening a document folder.

5.2. Refinement through selection and the addition of views

This view is refined by the selection of a second view on 'Business and competition'. This reduces the number of documents from 14,302 to 2745. The user opens the folder to expand the business and competition view under 'types of undertaking' and then 'undertaking'. The user selects to refine by selecting all documents in 'undertaking' which are marked as seen in Figure 13.

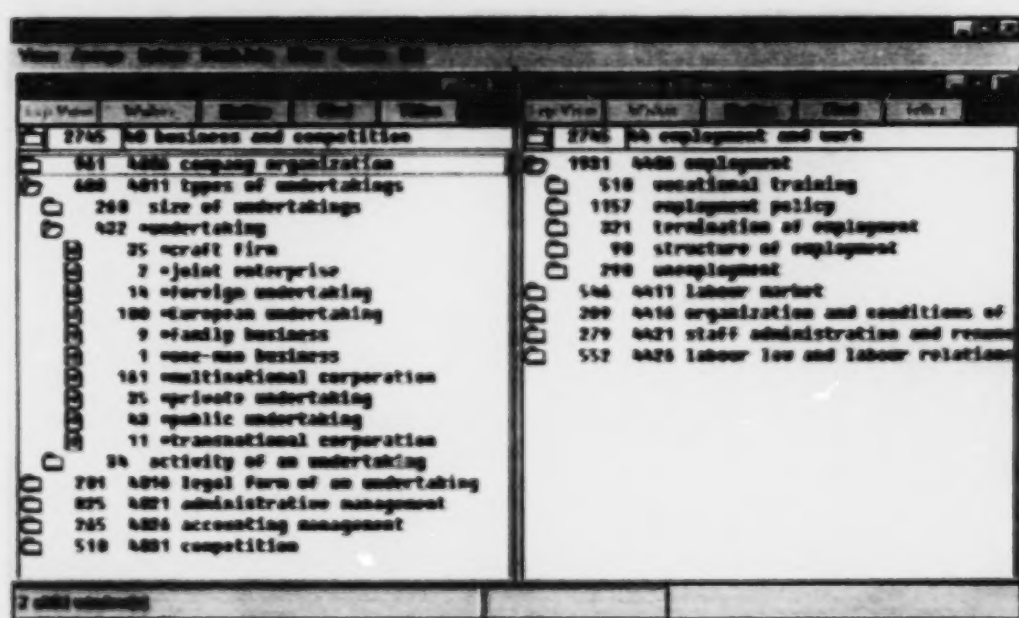


Figure 13: Combining 'Business and Competition' and 'Employment and Work'.

This clears the 'Business and competition' view of all except the selected documents and refines the view on 'Employment and work' as shown in Figure 14.

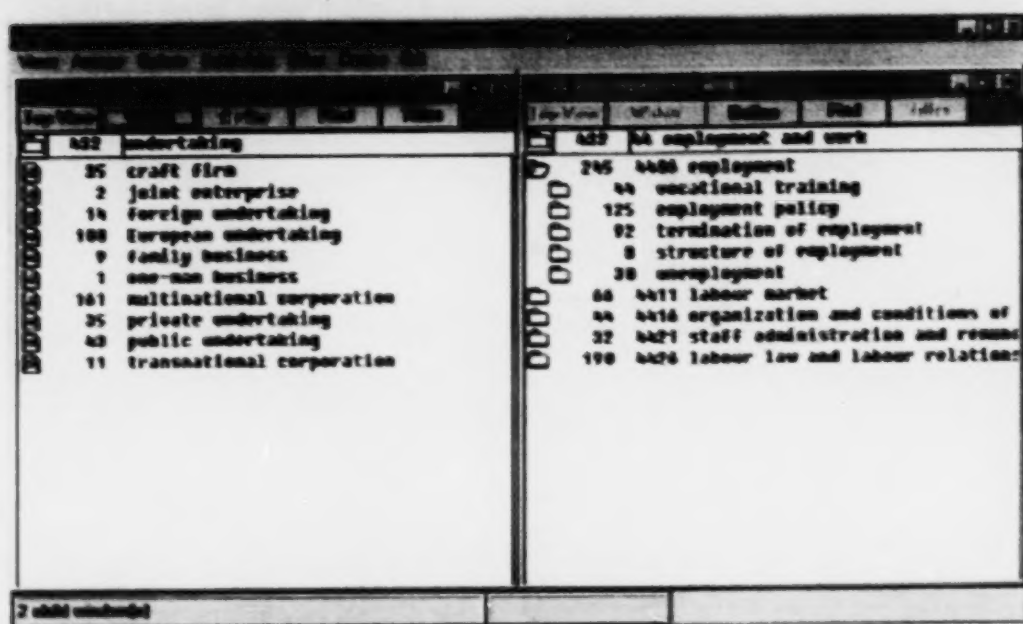


Figure 14: Refining 'Employment and work' by 'Undertaking'.

Additional refinements are pursued through the selection of labour law and labour relations and the addition of the top level industry view (Figure 15).

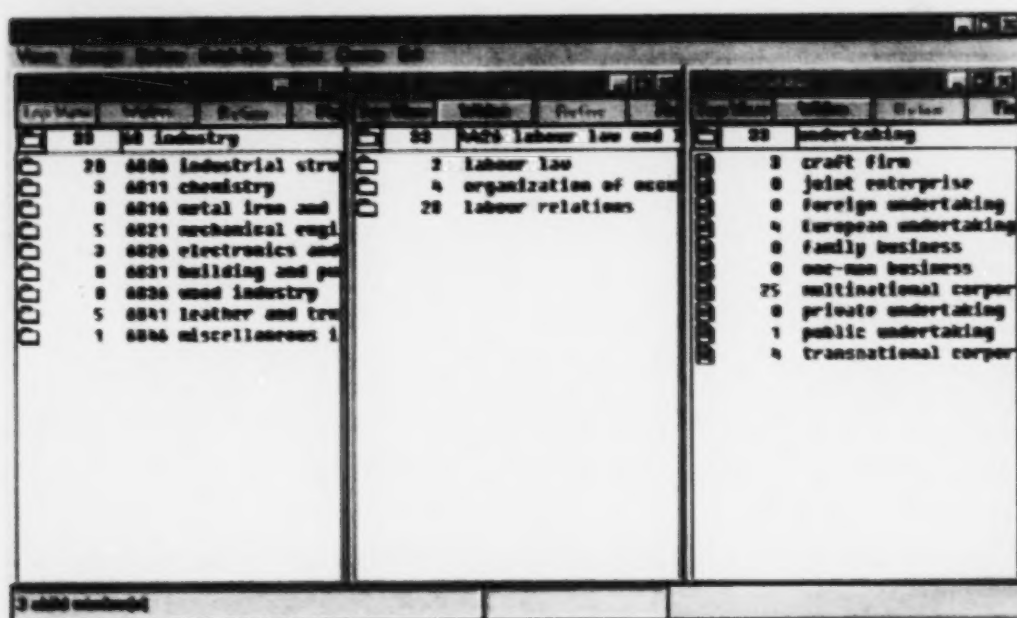


Figure 15: Additional refinement through selections and adding the view on Industry.

6. Conclusions

Effective disintermediation in information retrieval requires the development of tools and techniques that make it possible for the user to satisfy an information need without investing in the acquisition of skills and knowledge of the search process, currently the province of information scientists and librarians. The view-based searching techniques applied in the HIBROWSE prototype systems, described above, demonstrate how the incorporation of knowledge structures, primarily thesauri and classification schemes, in the user interface enables both detailed and general examination of the contents of a database and extends the scope for the user to control the interaction by adding, expanding, refining or removing views into the database. These systems also suggest how users can more effectively interact with a retrieval system when they are unable to specify the subject of their need, or can only express it in general terms and would like simply to explore the contents of the database.

The complexity and detail of indexing and retrieval operations can be hidden through the use of implicit Boolean searching, enacted whenever the perspective of a search is changed by an operation on just one of the views onto the database. Interaction with the system by the user is most concerned with the subject matter of the information need and the user is not concerned with the process of searching but rather with the specification of the subject matter of interest. Learning the capabilities of a system remains an essential for the professional intermediary through training courses, manuals and experience doing searching, but the capabilities of the system should ideally be self evident for the end-user and made more accessible, even if these are only required infrequently. The assumption that a sacrifice has to be made in performance where the interface is made more 'user-friendly' is being challenged through the application of techniques which emphasise the importance of recognition and selection and implicit searching as central to the design of online information retrieval interaction.

Where the view-based searching approach as implemented in HIBROWSE can be improved is the subject of on-going research; what is clear, by examining the workings of the current prototypes, is the increase in searching power made available without a commensurate increase in cognitive effort on the part of the user. Remembering commands and how to apply them is replaced by a limited set of actions which can reduce or expand a set of documents by refining or expanding, adding or removing views. The ease of use in World Wide Web navigation might now be complemented by an effectiveness in retrieval. Large result sets can be viewed according to helpful arrangements and divisions which indicate ways to limit or expand the amount of information being retrieved.

The absence of indexing using a controlled vocabulary for any information collection does not preclude the application of this approach. View-based techniques are applicable to both heavily indexed and free text databases alike, given there are suitable thesauri which, in the case of the latter, can be used to generate extended queries through the use of synonyms and narrower terms with stemming and other free text searching devices.

There will still be users who are content with the results of a 'quick and dirty' search, happy to use the five 'top ranked' documents from a set of many thousands, who will not reap the benefits of these efforts at disintermediation through the application of knowledge structures and associated search techniques. However, we trust there will be many others who, in order to exercise and extend their own professional skills and knowledge, will welcome the opportunity to discern the most apposite articles for the task they have in hand through the application of view-based searching.

Dr A. Steven Pollitt
Director, CeDAR — Centre for Database Access Research
School of Computing and Mathematics
University of Huddersfield
Queensgate
Huddersfield HD1 3DH
UK
Tel: +44 (0)1484 472147
Fax: +44 (0)1484 421106
E-mail: a.s.pollitt@hud.ac.uk

Acknowledgements

The HIBROWSE for Bibliographic Databases project and two of the authors are supported by a British Library Research and Innovation Centre grant RDD/G/282. Additional sponsorship from Sun Microsystems Inc., Software AG of the United Kingdom Limited, Elsevier Science BV, Elonex and the University of Huddersfield is gratefully acknowledged.

HIBROWSE is a registered trademark of the University of Huddersfield.

References

- [1] CRG (1955) *The Need for a Faceted Classification as the Basis of all Methods of Information Retrieval*. A memorandum of the Classification Research Group, published as UNESCO document 320/5515, 26 May 1955.
- [2] Davies, R. (1996) *Thesauri for the new century, Macrothesaurus no.4*, May, OECD Development Centre, Paris.
- [3] de Grolier, E. (1965) Current trends in theory and practice of classification, *Classification Research — Proceedings of the Second International Study Conference*, Elsinore, Denmark 14-18 September 1964, FID/CR Copenhagen.
- [4] Dix, A.J. (1991) *Formal Methods for Interactive Systems*, Academic Press, London.
- [5] Nicholas, D. (1995) Are information professionals really better online searchers than end-users? (And whose story do you believe?) *Proceedings of the 19th International Online Information Meeting* London, December 1995, *Learned Information*, pp. 383-397.
- [6] Occhini, G. (1983) Microelectronics and teleinformatics as economic and social disintermediation tools. (Elimination of intermediaries between producers, distributors and consumers), *BIAS Microelettronica. International Conference on Microelectronics and Teleinformatics*, Milan, Italy 23-25 February 1983, Federazione delle Associazioni Scientifiche e Tecniche, p. 9.
- [7] Payne, S.J. (1991) Display-based action at the user interface, *International Journal of Man-Machine Studies*, **35**, 275-289.
- [8] Pollitt, A.S. (1981) An expert system as an online search intermediary *Proceedings of the 5th International Online Information Meeting*, London, December 1981, *Learned Information*, pp. 25-32.
- [9] Pollitt, A.S. (1985) End-user Boolean searching on viewdata using numeric keypads, *Proceedings of the 9th International Online Information Meeting*, London, December 1985, *Learned Information*, pp. 373-380.
- [10] Pollitt, A.S. (1987) CANSEARCH: An expert systems approach to document retrieval, *Information Processing and Management*, **23**(2), 119-138.
- [11] Pollitt, A.S. (1988) A common query interface using MenUSE — A Menu-based User Search Engine, *Proceedings of the 12th International Online Information Meeting*, London, December 1988, *Learned Information*, Vol. 2, pp. 445-457.
- [12] Pollitt, A.S., G.P. Ellis, M.P. Smith, M.R. Gregory, C.S. Li and H. Zangenberg (1993) A common query interface for multilingual document retrieval from databases of the European Community Institutions, *Proceedings of the 17th International Online Information Meeting*, London, December 1993, *Learned Information*, pp. 47-61.
- [13] Pollitt, A.S., G.P. Ellis, M.P. Smith (1994) Peek-a-Boo revived — end-user searching of bibliographic databases using filtering views, *Proceedings of the 18th International Online Information Meeting*, London, December 1994, *Learned Information*, pp. 63-74.
- [14] Pollitt, A.S., G.P. Ellis, M.P. Smith (1994) HIBROWSE for bibliographic databases, *Journal of Information Science*, **20**(6), 413-426.
- [15] Ranganathan, S.R. (1965) Library classification through the century, *Classification Research — Proceedings of the Second International Study Conference*, Elsinore, Denmark, 14-18 September 1964, FID/CR Copenhagen.
- [16] Treglown, M. (1996) Results of applying the cognitive walkthrough method to the HIBROWSE for EMBASE system, *Research Report RR96/02*, School of Computing and Mathematics, University of Huddersfield.
- [17] Treglown, M. (to appear) Are database thesauri menus? To appear in the *Adjunct Proceedings of HCI '96*, held London, 20-23 August 1996.



Web page authoring tools: comparison and trends

Linda Craney

Prometheus, UK

Abstract: Software tools to author World Wide Web pages have evolved rapidly over the past year. Initially available from universities and individual enthusiasts, they are maturing into very feature rich applications and are now offered by large corporations. These applications are enabling more companies to create and maintain pages themselves on the Web or on corporate Intranets. The market continues to grow quickly as the Web matures. Looking at trends evolving amongst the products may assist purchasers to choose tools that fit their environment.

Keywords: HTML editors, Web authoring tools

1. Introduction

World Wide Web pages are composed in a language called Hypertext Markup Language or HTML. Actually, HTML is not a programming language at all. It consists of codes or *tags* which are placed in an ASCII text file. Software applications called *browsers* request these files from computers on the Web known as *Web servers*. The browser displays these text files as formatted Web pages according to HTML tags it reads. Tags instruct the browser how to highlight or align text, where to insert graphics, etc.

For example, words to be displayed in bold lettering would be written in an HTML file as such:

`<B>some important words</B>`

As the browser reads the `<B>` tag, it displays text in bold lettering until it reaches the closing tag `</B>`. The words would be displayed on a Web page like this:

some important words

Using standard ASCII text files enables the World Wide Web to support a heterogeneous population of computer platforms. Because the format is not proprietary to one type of computer, both Web servers hosting these HTML files and user computers, or *clients*, requesting Web pages may be Unix workstations, PCs, Macintoshes, etc. A browser such as Netscape Navigator or Microsoft Explorer adjusts the display for a Web page according to the type of computer and screen resolution on which it is running.

Web sites have been increasing at an exponential rate since 1993. A fast growing percentage of sites were published by companies during 1995. A survey by Matthew Gray at MIT shows that 18% of sites on the Web were commercial ones in December 1994. By January 1996, this percentage increased to 50%.¹ As companies used the Web to market their products and sell online services, the emphasis on Web page design and functionality increased. Netscape introduced HTML tag extensions to the HTML standard. Microsoft also introduced a browser and its own HTML tag extensions.

Text editors such as Notepad or SimpleText were among the few alternatives available for creating HTML files until 1995. Web page authoring tools or HTML editors began to appear from universities and individual enthusiasts. Copies were initially available to download for no charge. They looked and worked like simple word processors. Instead of typing HTML tags such as `<B>` into a text file, writers could select a tag by clicking on a button from a toolbar. Initial benefits from early HTML editors were that they provided basic editing functions and writers no longer had to refer to growing lists of tags.

Throughout 1995, these HTML editing tools grew in number and functionality. The market began to attract large corporations including Microsoft and Adobe. WYSIWYG ('what you see is what you get') interfaces appeared. Graphics editors began to be included into the applications. As Web sites grew in size and complexity, site management tools were also developed to assist authors in tracking page links and images.

It is difficult to estimate the number of applications available for authoring Web pages. A search in Yahoo yields 120 matches. This paper reviews five authoring tools that have received a great deal of media attention, awards and endorsements over the past year. Moreover, these products have remained current with new HTML tag extensions and their manufacturers seem committed to continuing product development.

This paper identifies and discusses trends that are evolving in the development of these authoring tools. It then reviews each of the tools and producing companies individually. Because the Web is evolving so rapidly these products are also being frequently upgraded. It would be misleading to provide a feature-by-feature comparison in a paper five months before publication. Some of these products have new releases pending in the next few months. Instead, this paper focuses on the products' overall design, ease of use and market position.

(Note — A table comparing individual features and example screen shots will be prepared from current software versions available in November and distributed at the conference.)

2. Authoring tools reviewed

Five software packages were reviewed here:

- Hotdog Professional from Sausage Software;
- HoTMetaL PRO from SoftQuad;
- PageMill from Adobe;
- FrontPage from Microsoft;
- Navigator Gold from Netscape.

3. Trends

The main trend evolving throughout HTML editors is that they are becoming WYSIWYG. This style hides the HTML tags from a user. Page layout is simulated to look like it was displayed by a browser. Users typically find this an easier way to author Web pages, although experienced authors often prefer working with text based HTML files. Hotdog is the only software reviewed that is not WYSIWYG. HoTMetaL continues to display files in a text based HTML format in its latest version. However, it incorporates some WYSIWYG features. Text is shown to reflect font size and styles and images are displayed.

Their value is increasing. Prices are dropping and more functionality is being added for little, or no price increases. Most dramatically, FrontPage price dropped from £463² in December 1995 to £99 and £73 for Office 95 users in June 1996 when it was released by Microsoft. This offer is available until March 1997. Overall these editors range in price from £53 to £106. Navigator Gold can be purchased for the least amount of £53 which includes the popular Navigator browser. However, the price increases up to £66 if someone also purchases a handbook and £33 more for a year's subscription to upgrades. PageMill and Hotdog cost £66. HoTMetaL version 3.0 costs £106, up from £99 for version 2.0.

Increased functionality in the products includes easy colour selection and image manipulation. Until recently, working with images was cumbersome. Different tools or software applications were required to make backgrounds transparent or change the image format to .GIF or .JPEG. For PC users, some of these tools were only available as MS-DOS routines. SoftQuad has integrated its MetalWorks graphics editor into version 3.0 of HoTMetaL. HoTMetaL, PageMill and FrontPage convert graphics to a .GIF format automatically as the image is inserted into a page. Making the background transparent is also becoming as simple as a couple of mouse clicks. Selecting colours from colour wheels or RGB palettes now replaces entering hexadecimal codes to adjust shades of red, green and blue for a colour.

Incorporating programming functions such as forms handling is being introduced. FrontPage includes a proprietary set of routines called *bots* which allow forms handling and search capabilities. HoTMetaL includes sample CGI scripts. Sausage Software offers add-on products called *Snaglets* which include Java functionality for animating images.

Site management is becoming more developed in the product offerings. FrontPage integrates a sophisticated site management tool in the product. Adobe offers similar functionality through a sister product to PageMill called SiteMill which costs £200 for registered owners of PageMill. Hotdog allows users to group files into Web projects which may be opened as a group. A report can then be run to show a listing of page links and images. Another handy feature this product provides is the ability to search-and-replace throughout multiple files.

All of the tools share a basic look and functionality. They offer the user a toolbar of buttons and pull down menus representing HTML tags. Paragraph or Normal is typically the default style that produces normal text. A user then selects a style such as Heading Size 1. All text typed until the style is changed will be displayed in a large heading size.

Context sensitive controls is another popular feature. HTML tags have a growing list of *attributes* or parameters available which allow authors more functionality to adjust the tag's display. When an author uses a particular HTML tag in an editor with context sensitive control, an associated dialogue box displays available attributes for the HTML tag. PageMill provides this functionality most easily through a floating dialogue box called the Attributes Inspector. Element attributes are available by clicking a button. Others like Navigator Gold and HoTMetaL have dialogue boxes that are available from a pop-up menu displayed by clicking the right mouse button.

Applications are growing larger and hungrier for memory. They are moving towards supporting, or in some cases requiring, 32-bit TCP/IP stacks. FrontPage and Navigator Gold are only available for Windows 95 or NT.

Online Information 96 Proceedings

Page 448

Hotdog offers a 32-bit version as well as a 16-bit version. FrontPage also requires 16 MB of memory to run its Personal Web Server. Sausage Software also recommends 16 MB of memory to run its 32-bit version of Hotdog as well.

The PC platform is attracting more attention from these manufacturers than Macintosh or UNIX. Though HoTMetaL is available on the widest range of platforms including PCs, Macintoshes and UNIX workstations, version 3.0 is only available for PCs including Windows 3.xx, 95 and NT. Adobe PageMill is currently only available on the Macintosh. In August 1996, version 2.0 will also be available for Windows.

4. Individual product reviews

4.1. Hotdog Professional from Sausage Software. Version 2.0

Hotdog is an entrepreneur's dream story of how the Internet can turn an idea into a multi-million pound a year business. First introduced in a newsgroup in the Spring of 1995, Hotdog and other add-on products are forecast to generate almost £3.5 million of revenue for Sausage Software of Australia. Product sales and advertising were solely over the Internet until June 1996.

The company was also the first to announce a Java tool for users interested in animating pages without learning the Java programming language. The product, Egor, steps users through creating an animation of a graphic. The company strategy is to continue development of niche Java applets and application software tools called Snaglets.

Hotdog is not a WYSIWYG editor. HTML is displayed in text based form. Tags are displayed in a different colour from text, though, which makes viewing or troubleshooting easier. Tags may be inserted by clicking buttons, pull down menus or from an alphabetical listing which a user can display on top of the Web page he or she is authoring.

Hotdog is very customisable. A user has a great deal of flexibility and control in using the software. Another distinguishing feature of the editor is its file handling. Files are automatically saved, easily converted from PC to UNIX format, and opened in tandem with others automatically as a project. Probably one of the most welcome features was the ability to search and replace text through multiple files.

Hotdog costs £66. The Snaglets are available separately or bundled with new versions of Hotdog. Support is available only by e-mail. There is no user manual but there are tutorials on Hotdog as well as HTML built into the product and extensive online help that is context sensitive. It is available for Windows and Windows 95. There is no information about plans to release a Macintosh version.

If someone is comfortable working with text based HTML, this is a good tool to use. Another tool is required for handling images. A free 30 day evaluation copy is available to download from <http://www.sausage.com>.

4.2. HoTMetaL PRO from SoftQuad. Version 3.0

HoTMetaL grows out of SoftQuad's expertise with Standard Generalized Markup Language which became popular during the 1980s as a standard for electronic document exchange. HTML is based on SGML. Like Hotdog, HoTMetaL displays a text based HTML file. The tags are displayed as small graphics which make them easy to distinguish from other text. Tags are inserted from pull-down menus or tool bars. A real-time syntax checker will not allow someone to enter an invalid tag.

HoTMetaL includes many WYSIWYG features. Font sizes and paragraph styles are displayed as WYSIWYG. Hypertext links are displayed as they would look through a browser. The package provides over twenty template pages including a commercial home page, Press Release and Product Descriptions examples. Word processing documents are easily converted in this application.

Images are converted automatically as they are inserted into a page. Colours are easily changed via a dialogue box. The package now integrates its MetaWorks graphics editor which was previously available separately. It provides the user with many features to manipulate images. Image mapping is also easily accomplished by dragging a shape tool over the desired section of the image. Image backgrounds can be created as easily as well by clicking on the colour. Adjusting image colour and other formatting features are available. Image effects such as making 3D buttons are also available.

The company plans to maintain development of multimedia support including Java and Microsoft's ActiveX standard. Sample CGI scripts are available but they are not integrated with the software.

The software is available for Windows 3.xx, 95 and NT. Version 2.0 is available on a wide range of platforms: Windows, Macintosh, Sun/SPARC, DEC Alpha, IBM RS6000, Silicon Graphics and HP9000/700. The manual is comprehensive and includes a tutorial geared towards teaching someone basic HTML. Telephone support is available for 90 days.

HoTMetaL is a good choice for a company looking for a wide range of tools within one package. A freeware copy of version 2.0 is available to download from the company's Web site, <http://www.sq.com>. HoTMetaL has an extensive range of features and provides easy interfaces for handling images and creating tables and frames. SoftQuad is serious about promoting its product in the UK and has been exhibiting at Internet and computer exhibitions. Its syntax checker may become frustrating for experienced authors who want to insert their own tags and alter the file layout.

4.3. PageMill from Adobe. Version 1.0

PageMill from Adobe was the first WYSIWYG Web authoring tool available. Produced by Adobe, it gives a user the creative power you would expect from an Adobe product while keeping technical aspects hidden.

First impression of this editor is its simple and minimal interface. This is due to one of its key features, a floating panel called the Attributes Inspector. This panel is context sensitive. It contains buttons to adjust page and text colours and styles. A third button allows users to view available attributes selected objects. For example, if a user selects a horizontal rule or bar across the page, the Attributes Inspector displays attributes available for the HTML tag `<HR>` such as adjusting the bar's width and height.

PageMill supports many intuitive drag and drop features. Elements can also be adjusted directly on the page by selecting them and then dragging a selection handle to adjust their size. Hypertext links are created by dragging a Page icon from the destination page to text on a source page.

Again, as one would expect from Adobe, working with images and colours is made very easy. Colours for text and background can be chosen from a colour palette or RGB wheel instead of using hexadecimal combinations for shades of red, green and blue. Creating a transparent .GIF is as simple as a couple of mouse clicks. Converting image formats for display on a Web page takes even less mouse clicks. None, in fact: PageMill does it automatically.

PageMill also allows users to insert text and HTML tags by working in a Raw HTML mode. This gives users access to HTML tag extensions the product is not yet supporting.

PageMill costs £66. Version 1.0 is only available for Macintosh but version 2.0 is scheduled to be available in August for both Macintosh and Windows. Upgrades are free. The product includes a manual and tutorial. Telephone support is available for 90 days. SiteMill, its sister product, is available for site management for PageMill owners at £200.

PageMill is a powerful tool that is very simple to use. The ability to include text based HTML tags provides authors with flexibility and insurance for new HTML tag extensions. Moreover, it will probably become tightly integrated with Adobe's PageMaker for desktop publishing and Acrobat for creating and viewing .PDF files. Like SoftQuad, Adobe can typically be found at Internet and computer exhibitions. This is a good way to see a product demonstration. Its pricing for SiteMill, its site management tool, will come under pressure from Microsoft's FrontPage release. FrontPage includes a site management tool, Explorer, in the application.

4.4. FrontPage from Microsoft. Version 1.1 Beta

Initial impression of this product is that it looks very much like Microsoft Office Products. This may not sound so surprising except for the fact that Microsoft did not develop FrontPage. It was developed by Vermeer Technologies from Cambridge Massachusetts in 1995. Vermeer was purchased by Microsoft in 1996 for a rumoured £87 million, according to *Internet World* magazine. Representing another Cinderella Internet story, the company began with £17 million of capital earlier that year.

FrontPage is a WYSIWYG editor. A user can easily view the HTML source for a Web page but cannot modify it from within the software. The product supports standard Windows keyboard shortcuts. For example, typing CTRL+HOME will position the screen display at the top of the page. Moreover, Office documents can be accessed in an internal site but not from a public Web server.

Working with images is simple and rich with features. Images are converted automatically if they are in an incorrect format. Image maps are created by dragging a rectangle over the desired area the designer want to make a hotspot. Transparent backgrounds are created instantly by simply clicking a tool and then clicking the area of the image to become transparent. Interlacing can be added by clicking a checkbox in a dialogue box. Colours can be changed from a dialogue box using colour palettes.

Unlike other products, FrontPage has its own jargon. A Web site isn't a Web site in FrontPage, it is a *Web*. Anchors are called *bookmarks*. It introduces the concept of a *bot*, or *WebBOT* as it is trademarked. Bots are robots or routines which allow authors to add sophisticated functionality that would typically require programming. Features include searching through a Web Site and adding form registration results to an easy-to-read file. A unique feature of Front Page is its *Include bots* which allow authors to use the same copy of text or portions of pages throughout their Web sites. When the original is modified, all of the Web pages using the text are also updated automatically.

Using *bots* requires that FrontPage extensions or programming code must be stored on the author's Web server. This may be only an administrative issue for corporations maintaining their own Windows NT servers. Others, however, must rely on their Internet provider to install the extensions for them where their Web pages are hosted. Most providers in the UK use UNIX based servers, not Microsoft NT. A survey of the top ten UK Internet providers from *Internet* magazine's Top of the Pops listing revealed that no provider had immediate plans to implement the extensions but several were reviewing the possibility. Non-UNIX support and potential system conflicts were cited as issues.

(Note — an updated and more comprehensive survey of Internet providers will be made in November and distributed at the conference. FrontPage is just being released by Microsoft as this paper is being completed.)

FrontPage was released as a Microsoft product in June 1996. It is only available for Windows 95 and Windows NT with their 32-bit TCP/IP stacks. Software will be available at software retail stores and mail-order. A site management tool called Explorer and server software come bundled with the editor. It is worth noting this is a large application to install requiring 15 MB of disk space. Price is £99 or £73 for Office 95 users until March 1997. The tutorial with the product is informative and easy-to-follow. A manual will be available. Telephone support is

Online Information 96 Proceedings

Page 450

available from Microsoft for 90 days.

FrontPage provides authors with a rich environment to create and manage Web pages. Microsoft users will find it easy to use and thus it will proliferate quickly through companies already standardised on Microsoft products. A handful of Internet providers will begin to implement the server extensions, viewing this as an opportunity to differentiate themselves amongst the 150 Internet providers in the UK today. Because FrontPage is a very extensive application which includes a personal Web server, it may prove challenging for a technical novice to install and maintain.

4.5. Navigator Gold from Netscape. Version 2.01 Gold

The Navigator Gold browser includes a WYSIWYG Web authoring tool. Web pages are created or modified by selecting editing commands from the File menu. Unlike other WYSIWYG editors, Navigator Gold allows users to modify the source text and HTML tags for a Web page. Like other authoring tools, toolbars are available to manipulate text and page styles. One contains a useful button for increasing and decreasing font size. A context sensitive dialogue box for modifying attributes for a selected element is available from a pop-up menu. This is available by clicking the right mouse button.

Images can be inserted if they are .GIF or .JPEGs but not edited within the software. Navigator Gold does not provide an image editor but allows users to link to an application of their choice. Changing link and background colours is simple. A Document Properties dialogue box allows individual colour selections. It also provides authors with a selection of predesigned colour schemes where the background colour and link colours have been co-ordinated. The ability to preview the colours in the dialogue box itself is also helpful.

Netscape provides a rich source of templates and samples from its Web Site. These pages include Web page templates ranging from family to business pages, clip art, backgrounds and links to Java scripting resources. Template pages may be viewed from the site and then downloaded. Creating Web Services under the Help menu is also a good resource for information on using HTML tags, tools and design tips. Netscape also provides users with wizards to help them create Web pages. They are interactive pages that work from the Netscape Web site.

Navigator Gold is available from Netscape's Web site or through its resellers. The browser and Web authoring tool prices begin at £53 for a software license only. This does not include a manual or upgrades to new releases. Software with a manual costs £66. A year's subscription service to upgrades costs an additional £33. The manual assumes the user is a novice and provides an introduction to HTML and related terminology. The product is supported by Netscape's online support centre which provides assistance via a Help Request Form and telephone support for 90 days.

Navigator Gold can be an easy and inexpensive way for someone to start creating because it is included with the browser. However, other applications offer more functionality for working with images. This may be addressed in a pending upgrade. Accessing Help information and other supportive features from the Netscape Web site is cumbersome compared to other editors where supporting tutorials, wizards and Help information are included and run locally.

5. Conclusions

Conclusions are that editors will evolve into very feature rich applications that are easy for non-technical authors to use. Future tools will have a WYSIWYG interface which will mask the need for users to know HTML tags. For most Web pages authors this is good news. Experienced Webmasters will continue to prefer the ability to work with HTML tags. Simple, non-programming interfaces to functions such as searching and form registration will become more developed and easy for non-technical authors to use. Image handling is now mature enough in most products to satisfy commercial users but those seeking high design capabilities will require other image manipulation programs such as Adobe Photoshop.

Enhancements and new releases will continue to be introduced frequently. Independent producers will continue to support both Microsoft and Netscape HTML tag extensions. It is more difficult to determine what Microsoft and Netscape will continue to propose and support. Netscape enjoys an incumbent advantage being the first commercially available browser and capturing 83% of the market, according to the company.³ Microsoft's FrontPage programming extensions only work with its NT operating system, not UNIX. For a company maintaining its own Web server, this may not be an issue. However, for companies and individuals who use Internet providers or hosting services for their Web sites, this may be a problem because most of these organisations use UNIX.

Throughout 1997, users may want to consider having a combination of tools available to produce and maintain their Web pages. For example, a WYSIWYG editor can be a useful way to begin designing and authoring a series of Web pages but simple maintenance may be faster throughout the next year via a text based HTML editor. Evaluation copies and freeware software represent a good way for users to introduce themselves to Web Authoring tools. Hotdog Professional is available for a 30 day evaluation and HoTMetaL version 2 is available for no charge.

Linda Craney
Prometheus
14 Noel Road
London N1 8HA
UK
Tel: +44 (0)171 704 8700
Fax: +44 (0)171 919 6380
E-mail: linda@prometheus.co.uk
<http://www.prometheus.co.uk>

Notes

- (1) Internet Statistics: Web Growth, Internet Growth. Andrew Gray.
<http://www.mit.edu:8001/people/mkgray/net/web-growth-summary.html>.
Commercial Web sites are shown as URLs with suffix .com.
- (2) These products are priced in US dollars. Prices were converted to sterling using a rate of 1.5 dollars/pound for UK readers' convenience.
- (3) Netscape [Client Products]. http://home.netscape.com/comprod/partner_central/products/client.



How can documents be put onto the Web?

Ross Dixon
Pindar PLC, UK

Abstract: *The World Wide Web is rapidly maturing as a medium for publishing documents online. This makes it important to both publishers and other organisations with an online publishing requirement. The specific characteristics of the Web mean that publishing using this medium has a number of unique requirements and presents a different challenge to anyone wishing to put documents on the Web. Key issues which have to be resolved include determining the nature of the documents to be published, identifying whether documents should be published directly or from an intermediate form, and deciding what sort of document creation tool should be used. All of these issues are addressed within this paper explicitly and with the use of case studies.*

Keywords: Internet, Web, documents, Acrobat, HTML

1. Introduction

The World Wide Web, or simply the Web, has grown out of the Internet and become a significant publishing medium. Estimates of beyond 50 million users have been made and many organisations (publishers and others) have identified the Web as being integral to their approach to electronic publishing. Clearly the Web has many applications and most Web sites are aimed at marketing rather than serious publishing. However as the understanding, size and infrastructure of the Web continues to increase then scope for its use as an online publishing medium will continue to grow. The more recent intranet phenomenon aimed at internal publishing within organisations further moves the emphasis from marketing to serious publishing use of the Web.

Already the Web has a number of significant publishing sites (many of them in their pilot stages). The experience gained by Web publishers shows that various pitfalls exist for the potential Web publisher. This paper aims to define a brief route map for anyone wanting to get their documents onto the Web and is based on experience from a variety of projects.

2. What types of documents can be published on the Web?

A key decision for any Web publisher is the form of the documents to be published. While many forms of document can be published these can be categorised in three ways:

2.1. Formatted documents

These cannot be edited or have their contents searched. A typical example of such documents are images, i.e. TIFF images. The rather unintelligent nature of such documents is well known and not well suited to the Web. In particular their requirement for indexing (no content searching), no scope for hypertext linking and their data intensive nature (i.e. slow to download) are particularly problematic. For these reasons, this type of document is less common on the Web and is not considered further in this paper.

2.2. Processable documents

The content of these documents can be processed, i.e. searched and/or edited. Processable documents are what we typically think of when we talk of electronic documents — text (i.e. from a word processor) is the obvious example. In comparison with formatted documents, processable documents appear to be much more suited to the Web, i.e. they are content searchable, allowing hypertext linking and less data intensive. An often overlooked characteristic of processable documents is that they do not have to exist as a collection of pages and this non-page approach is supported by the Web. In fact these types of documents are by far the most common on the Web and a specific approach to 'marking-up' the text for the Web known as HTML (Hyper Text Mark-up Language) is the commonest approach to text publishing on the Web.

2.3. Formatted processable documents

This is a somewhat more unusual concept and can be thought of as freezing a processable document such that it is still content searchable and allows hypertext linking but usually does not allow editing and requires that documents do exist as a collection of pages. As expected, this hybrid approach has characteristics of the other two document types but most interestingly maintains most of the benefits of processable documents while forcing the familiar page approach to be retained. The most popular implementation of the formatted processable approach is Adobe Acrobat.

3. What is HTML (hypertext mark-up language)?

HTML is the native language of the Web. The HTML approach requires that document contents are tagged (marked-up) with codes to denote the document presentation and structure. This allows the marked-up contents to be presented and displayed by a Web browser according to the intended style of the Web publisher. In addition to arranging text as required, HTML codes can also be used to make hyperlinks (connecting together non-contemporaneous document sections or different documents), creating forms and referencing graphic data. The benefit of this approach is that the publisher has great flexibility regarding what can be published and has absolute control over the published material. Most important of all, since HTML is the native language of the Web, all Web browsers understand (some versions of) HTML and display HTML pages. Largely because of these characteristics, the vast majority of Web pages are HTML pages.

4. How are HTML documents created?

Three methods of generating HTML documents are relevant and are described below:

4.1. HTML documents

These can be created by any word processor because the HTML codes (i.e. the tags) used are defined using the standard ASCII character set which can be input directly from the keyboard into any word processor document. While this may appear to be the simplest approach (making use of existing applications), it is not commonly used because it requires a high degree of HTML expertise from authors and in comparison with the other approaches is cumbersome and error prone.

4.2. HTML editing packages

Specific HTML editing packages exist which eliminate much of the tedium associated with creating HTML pages directly from a word processor. HTML editors may be associated with word processors or other applications such that the editor is an add-on application which 'HTML-enables' the word processor. This approach has the benefit of familiarity for the word processor user and is therefore suitable for casual and low volume HTML document production. However, this approach is limited by the characteristics of the word processor which was almost certainly not designed for HTML output. The alternative approach for higher volume HTML document creation is to use a dedicated HTML editor which is not associated with a word processor.

4.3. Generate HTML files automatically

The final approach is to generate HTML automatically from other data files. Once again two methods are possible here: the first is to use a document creation package such as a word processor or desktop publisher which can export documents to HTML form. At present many such packages do not have these facilities and those that do suffer from the same problems as the add-on editors describe above. The second method is to export HTML data from SGML (standard generalised mark-up language) data. SGML is the prevalent data format used in electronic publishing and is closely related to HTML. In fact SGML inspired the creation of HTML and HTML is often described as a sub-set of SGML. The benefit of this second method is that data in SGML form is relatively easy to export to HTML, can be used in non-Web publishing applications (i.e. to CD-ROM or paper) and benefits from a relatively mature infrastructure which has developed in the 10 years since SGML was formally standardised.

5. What is Adobe Acrobat?

Adobe Acrobat is aimed at allowing users to distribute, interchange, view, annotate and print most forms of electronic document, i.e. a text file from a word processor, a spreadsheet file, DTP graphics and document images. This approach involves converting documents to a proprietary form which can be viewed by a viewer running on most desktop computing platforms. While Acrobat is the most popular approach in this class it is worth noting that comparable products exist, notably Envoy (from Corel) and World View (from Interleaf). For the sake of brevity only Acrobat is discussed in detail in this paper.

Acrobat (and other document viewers) aims to overcome the traditional problems of electronic document

interchange such as Word Perfect users attempting to read documents from various versions of MS Word. Historically the only approaches available here were to standardise on application packages or to use an interchange format. The problem here is that enforcing standardisation is difficult even within an organisation and across different organisations it is impossible. Also the main interchange formats are restricted to specific data types, i.e. RTF for text, IGES for CAD drawings, TIFF for document images and so on. The Acrobat approach is a development of the interchange format which allows practically all electronic documents to be held in a single format which can be viewed using a single viewer.

So why is this important? Well, the Acrobat approach (remembering that other products offer similar functionality) allows distribution and interchange of electronic documents across most computing platforms without getting bogged down in compatibility issues.

6. How are Acrobat documents created?

A cornerstone of Acrobat is that any document which can be printed to a PostScript printer (most printers fall into this category) can be input into the Acrobat format known as PDF (Portable Data File). Two methods of creating the PDF from the electronic document are possible:

- PDF output can be generated directly from the document creation application (such as a word processor). This will usually involve the use of an additional Acrobat package known as PDF Writer;
- PostScript data (generated by the document creation package) can be converted to PDF by the use of Acrobat Distiller.

Both the above approaches exclude paper documents. A further package (known as Acrobat Capture) caters for this situation and is described later. One of the key features of the Acrobat approach is that specific software (from Adobe) is required and in this sense it is not considered as 'open' as the HTML approach.

7. Key characteristics of Acrobat/HTML and to what Web applications are each suited?

Acrobat PDF maintains the concept of pages and does not allow editing of contained data — it does allow simple re-use of this data (cut & paste) and annotation of pages. Graphics are included as part of the PDF file. This makes Acrobat a useful form for page publishing on the Web.

HTML does not require the page metaphor to be maintained and allows contained data to be edited for re-use and repurposing. HTML documents can refer to other data such as graphics which are separate from the HTML document. This makes HTML suited to most forms of Web publishing and in particular where complicated documents with various document components are involved.

8. What if your documents only exist in paper form?

It is possible to convert paper documents directly to either PDF or HTML forms, or to convert to SGML and then output to PDF or HTML. For all of these approaches three general stages are required (unless documents are to be manually re-keyed):

8.1. Conversion from paper to image form

This activity simply involves scanning paper documents to produce (TIFF) images.

8.2. Conversion from image to text

This is a character recognition process which usually involves quality assurance and correction/editing processes. If conversion direct to PDF form is required then the recognition process can be undertaken within an Acrobat application known as Capture.

A key point about this process is that many recognition systems do not have a good concept of page layout — they can recognise collections of characters but for example are not good at recognising columns. While this is catered for within some packages such as Capture (although these packages tend to take a considerable time to recognise layout features), other packages may require some manual intervention here if the layout information of the document is to be maintained.

8.3. Mark up/output

From the recognition output the HTML or SGML tags can be inserted (semi-automatically or manually). In the case of Acrobat, multiple PDF files can be collected together, hyperlinks and index data added.

Experience shows that conversion from paper adds a significant cost and time overhead to the process and wherever possible use of already existing electronic files should be made.

9. Case studies

Case studies of the production of HTML documents or Acrobat documents are very common given the large uptake of these techniques: however conversion direct to Acrobat from Paper or the use of SGML as a basis for output are more recent developments and case studies are less common. Two such projects with which Pindar has been involved are an Acrobat Capture project for Oxford University Press and an online and paper publishing from SGML project for the Institution of Electrical Engineers. These are described briefly below:

9.1. Acrobat Capture case study

Earlier this year Oxford University Press (OUP) identified two separate pilot projects in order to assess the viability of Acrobat as a publishing format and Capture as a tool to make documents available in this form. The two projects related to separate publications. The first is a collection of six discontinued books known as the Physics Monographs. The aim of this project is to allow Web access to a subset of each book such that potential users can identify whether the book is of interest and which sections should be retrieved. Access to these sections will require a subscription which will give the user the key to the encryption allowing access. Many aspects of online electronic publishing are covered in this project but the key issue for Capture is that a discontinued book was brought back to life without involving new print runs and so on.

The second project related to a single book, *Leukocytes*. This is not to be made available over the Web (CD-ROM only) and is therefore not considered further here.

The capture from paper to image form was a fairly simple task because of the relatively small volumes involved, i.e. 3000 pages for the Monographs.

The recognition of the characters from the images was more problematic. This was because of the very processor intensive nature of this task and that at the time of the project the Capture product was still at version 1 and had a number of bugs. Most notable was a practical limit on the number of images which could be batched for recognition. This recognition task is ideally left as a batch process to run over night. However the version in use at the time would not allow more than 140 images to be batched for recognition. This severely limited the scope for batch processing and made the recognition process very slow and expensive.

After the recognition process had been completed the correction stage was undertaken and this took considerable resources for various reasons. Most important were the high standard required by OUP, the relative complexity of the books and a limitation within Capture where much of the text formatting data (bold, italics and so on) is not picked up by the recognition engine or the automated error detection, so this has to be identified and corrected manually. Finally the equivalent of a Contents Page is created within the PDF document by the addition of bookmarks which denote sections within books.

Access to the bookmarks of the Monographs will be freely available on the Web and also to all index, reference, contents, bibliography and symbol pages of the books. The idea here is that users will be able to browse this information and then decide whether they want access to the body of the books for which they will have to take out a subscription.

9.2. SGML case study

The Institution of Electrical Engineers (IEE) publishes several learned journals. Pindar is responsible for the production of two of these, presenting them on paper and making them available online through the Online Computer Library Center (OCLC).

The two journals are *Electronics Letters*, published 25 times a year and *IEE Proceedings*, published six times a year under 11 titles, giving 66 publications annually.

IEE chose SGML as the natural format to provide the base structure of its documents. However it was confronted with the problem that SGML encoding of the text and producing the typeset pages had until then been viewed as two separate processes. If these processes were distinct it would mean two things — firstly that the data would have to be processed twice, at a great cost to IEE, and secondly that there would be a weak relationship between the SGML and the data on the typeset page. It is important to IEE that the data in the electronic version is exactly the same as that in the printed version to ensure a consistent publication.

The approach taken to address this requirement is that articles from IEE are received in paper form, and are scanned to image form and then converted to text using character recognition. The recognition output is manually cleaned and marked-up. Any equations are converted to TeX form and held as EPS (Encapsulated PostScript) files: figures are maintained as TIFF images. The EPS and TIFF files are embedded into the articles as images.

Once the SGML data has been produced the different outputs are generated. For hard copy, the SGML data is imported into a DTP application which formats the articles according to the IEE typesetting specification from which film and printing activities are undertaken. The OCLC online database is populated with the SGML data and finally PDFs are produced for OCLC to provide document output for subscribers.

10. Conclusion

Very wide scope exists for making use of the Web as a publishing medium. As the relevant infrastructure widens and improves it is anticipated that Web publishing will become increasingly important. A result of this is that the serious Web publisher will be confronted with the practical issue of how documents can be converted to a

suitable form. In the short term this can be addressed by the use of existing packages for the production of HTML documents. Much more problematic is dealing with documents which only exist in paper form and ensuring that electronic documents are 'future-proofed' for unforeseen versions of HTML or non-Web media. Fortunately the electronic publishing world existed long before the Web and familiar techniques such as document conversion and mark-up provide a credible approach to addressing these issues despite their recent emergence.

Ross Dixon
Pindar PLC
Ryedale Building
60 Piccadilly
York
YO1 1NX
UK
Tel: +44(0)1904 613040
Fax: +44(0)1904 613110
E-mail: R.Dixon@pindar.co.uk



Added value in electronic publications

Theo J. D. Bothma

University of Pretoria, South Africa

Abstract: Electronic publications are flooding the market, be it on diskette, CD-ROM or the Internet (in addition to the many informal publications on the Internet). Some of these publications are created specifically for the electronic environment, but many are conversions of existing material to electronic format.

It is not worth the time and effort merely to publish existing material in electronic format if no value is added in the conversion process. The processes by means of which value can be added include:

- the addition of various media such as graphics, sound, animation and video;
- the inclusion of tests, games and guided tours;
- the possibility of integrating a number of electronic publications;
- sophisticated access methods by means of hierarchical structuring, navigational aids and hyperlinking; and
- sophisticated full-text search facilities which assist the user with, *inter alia*, the use of Boolean operators.

The above-mentioned items can be grouped together as either related to content or related to the way in which the information has been structured. Publishers should furthermore take special care in developing a user-friendly interface, which is both aesthetically pleasing and functional in terms of visual aspects and which gives easy access to the information.

An electronic publication should be evaluated on various levels, viz. content, as well as technological aspects such as structural properties, the various methods provided to gain access to the information, and visual aspects. Meaningful integration of all these elements is essential for the development of a successful electronic publication.

Keywords: Electronic publishing, value adding, hypermedia, multimedia, full-text searching, document structuring.

1. Introduction

The general availability of computers to the public and the phenomenal increase in computing power on the desktop are revolutionising access to information. Applications software such as wordprocessing, spreadsheets, graphics editing and database applications is commonly used. With menu driven systems and graphical user interfaces, ease of use has increased tremendously. 'The growing power of computers makes it easier to "hide" the true complexity of operations from users' (Ref 1). Computer literacy amongst the general public is also increasing at an unprecedented rate and many people now have access to powerful microcomputers at the office as well as at home.

Distribution of information via CD-ROM is becoming commonplace. Even three or four years ago CD-ROM publications were mainly aimed at information specialists, and therefore included primarily reference or bibliographic databases with only a limited number of full-text databases. This situation is now rapidly changing in that many software houses and publishers are beginning to exploit the end-user market by publishing books, journals, etc. which are primarily aimed at the end-user on CD-ROM. Many of these publications are created specifically for the electronic environment, but many are conversions of existing paper-based publications to electronic formats.

The possibility of multimedia publishing has brought a great stimulus to this market. Proper lay-out, good quality graphics, etc. enable software houses and publishers to produce the equivalent of a coffee table book on CD-ROM, with a number of added facilities such as easier access to the information through full-text searching and hyperlinking.

Any user of these new products, be it end-user or information specialist, has to judge whether the electronic publication offers any features which are not available in a paper-based equivalent, i.e. whether any real value has been added in the conversion process from printed to electronic publication, or in the creation of a new electronic publication.

The traditional library environment should carefully take note of these new developments, especially since there is an increasing demand that these products be made available in libraries. Librarians should therefore be able to judge the usability of these products, and be able to make recommendations as to whether items should be bought for the library and how these products can be used. A new element of evaluation is therefore brought in: publications cannot be judged solely on their content and general appearance, but also on how effectively information technology has been employed in creating a publication that is superior or inferior to an equivalent paper-based publication. If libraries and librarians do not offer these publications and a knowledgeable judgement on their usability, this function will be taken over by outsiders, and the library is in great danger of being side-lined in the information process. This necessitates a change in the attitude of librarians and information professionals, as already indicated by Feldman (Ref 2):

'If libraries do not address changing patterns of information use, then others will. [...] More than ever before, imagination, insight, understanding of new media, and enthusiasm for the process of ongoing change are going to become essential ingredients to effective librarianship.'

In this paper a number of processes by means of which value can be added in electronic publications are discussed, and the discussion is illustrated with examples taken from a wide range of electronic publications.

There are various ways in which value can be added to an electronic publication, viz.:

- the addition of various media such as graphics, sound, animation and video;
- the inclusion of tests, games and guided tours;
- the possibility of integrating a number of electronic publications;
- sophisticated access methods by means of hierarchical structuring, navigational aids and hyperlinking; and
- sophisticated full-text search facilities which assist the user with, inter alia, the use of Boolean operators.

When existing information products (books, articles, dictionaries, etc.) are converted to electronic format, the developers should take care that the electronic publication is not only an electronic edition of the existing paper format but that value is added in the conversion process to make the electronic publication more useful than the paper original. When a new publication is written especially for the electronic environment, the same care should be taken to make full use of the functionalities available within this environment.

2. Value-adding through adding multimedia

Because of the processing abilities of computers and larger storage media such as CD-ROM, etc. the traditional alphanumeric data is increasingly being supplemented by graphics, animation, sound and video. Multimedia — the integration of the previously mentioned data types — is now generally available on the average desktop PC, and it is safe to say that computing is becoming multimedia.

This new possibility opens up the way for every publisher to add multimedia elements to electronic publications. In many cases this is done very effectively. However, in some cases the usability of the multimedia clips can be questioned and the addition of these clips seems to be no more than a gimmick. One very often gets the impression that the clips were added simply because they were available, or because the publishers wanted to show their ability to add the various media. This is especially the case in multimedia encyclopaedias and similar publications, where the choice of multimedia clips very often seems to be totally random and no motivation is offered for the inclusion of specific clips (or the exclusion of others).

Sound, animation and video are usually available 'on demand', i.e. the reader can click on a button to activate the clip. A control bar which enables the reader to pause, stop or rewind the clip and which should indicate his/her progress through the clip should be available. Very often there are unfortunately no control bar, no option to stop a clip and no indication of where the reader is in the clip. One sometimes also finds that publications or sections of publications are introduced with long audio, animation or video sequences before the reader is presented with the actual material where an active choice can be made as to which information will actually be consulted. These methods of presenting multimedia information are not to be encouraged — the reader needs to be in control of the information space at all times.

2.1. Graphics and photo quality images

Graphics and photo quality images have been standard in many paper-based publications for many years. These features have only recently become available in electronic publications and most multimedia publications now include numerous images to supplement the text material. This is mainly due to the higher quality monitors that can display true colour effectively, and compression techniques such as JPEG compression and fractal compression by means of which file sizes can be reduced dramatically.

High quality graphics can be found in all electronic encyclopaedias, many Microsoft publications in the Microsoft Home series, as well as numerous other publications. Graphics are reproduced in full colour with remarkable detail and are equivalent to any printed in paper-based publications. One of the best examples of photo quality images is *Microsoft Art Gallery*, which contains the full collection of the National Gallery in London.

2.2. Audio clips

Audio clips are used in many different ways to enhance electronic publications. Unfortunately the addition of these clips seems arbitrary in many cases, especially in encyclopaedias, where the authors have often added a small random selection of sound clips, such as a few lines from a well-known speech by a politician, short musical clips from well-known works of composers, animal sounds, etc. These clips are usually so brief that they have extremely limited value.

Sound may also be totally useless in terms of the information it conveys in cases where it functions only as a background feature. In some of the Microsoft Home publications, for example, sound is very often played in the background to create an eerie atmosphere without the sound having any real bearing on the information being presented on screen: it is merely 'entertaining' (or could even be distracting to the serious reader). In these publications there fortunately is an option to turn the sound off.

In many other cases, however, sound is used very effectively.

Some dictionaries, for example, have a sound button which can be activated so that the word is pronounced; this is the case in the *American Heritage Dictionary* which forms part of *Microsoft Bookshelf*, as well as *Macmillan's Dictionary for Children* and *Dorling Kinderley's My First Incredible Amazing Dictionary*. In the *Dictionary for Children* sound is also used to explain the meaning of words; for instance, in the case of 'applause' a standard textual definition is given, together with a button which may be activated to let the reader hear what 'applause' sounds like.

In story books for children, for example the *Living Books* series, sound is also used very effectively (see below).

The best examples of the use of sound can be found in *Microsoft Musical Instruments*, a number of publications explaining great classical works (such as *Microsoft Multimedia Beethoven*, *Microsoft Mozart*, *Microsoft Stravinsky* and *Microsoft Schubert*) and some language-teaching CD-ROM publications.

Microsoft Musical Instruments, for example, contains descriptions of musical instruments from all over the world, together with sound clips of each of the instruments, as well as descriptions of various types of orchestra and accompanying sound files. There are usually at least two sound clips associated with each instrument, viz. a short passage played on the instrument as well as the range of individual notes of the instrument; the name of the instrument can also be pronounced. Figure 1 gives an example from *Microsoft Musical Instruments*, showing a single instrument with the 'sound box' activated; note the small buttons in the form of a speaker which can be clicked to listen to or stop the sound clip.

In the publications explaining classical masterpieces the accompanying text is richly interspersed with sound files. Sound files may either be midi files (synthesised sound) or direct links to the CD-Audio version of the music, with the reader often given the choice to hear either the midi file or the full orchestra. Sound can illustrate inter alia theme variations and the use of various instruments; often interesting points in the score are illustrated by giving the sheet music which can then be played in either midi format, or directly from the CD-Audio, as in Figure 2, taken from *Microsoft Multimedia Beethoven*. In Figure 3, also taken from *Microsoft Multimedia Beethoven*, the text scrolls on as the music, played from the CD-Audio, progresses; the words of Schiller's 'Ode an die Freude' change along with the music and the reader has the option of showing the words of the hymn in either German or English. The reader also has the option to pause the music at any time and to link to the 'Art of Listening'.

2.3. Animation and video

Animation can be used to illustrate any real life situation or process where video would provide too much detail. This is done very effectively in especially encyclopaedias such as *Microsoft Encarta*, *Compton's Interactive Encyclopedia*, the *New Grolier Multimedia Encyclopedia*, and so on.

Video is still fairly restricted in its application because of the computing power needed to decompress and show video on screen, as well as the size of video files. At present video is shown in eighth or quarter screen size windows, running at up to 25 or 30 frames per second, without the use of any additional hardware, in formats such as AVI (Video for Windows), Apple QuickTime and MPEG.

Especially animation and video files should have a standard control bar which allows the reader to play, pause or replay the video or animation sequence (as discussed above).

In the *Living Books* Series published by Broderbund (e.g. *Grandma and Me*, *Arthur's Teacher Trouble* and *The Tortoise and the Hare*) sound, graphics and animation have been combined to provide a brilliant learning environment for children. In each case it is possible to choose to have the text appear and read in English or Spanish (or even Japanese in some cases); the reader also has the option of only having the text read or to play inside the book. If the reader chooses to play inside the book, the active page is read, after which the reader can click on any word to hear the word again. Most graphic elements in the book are active, i.e. the reader can click on a graphic and it is animated together with a sound file. Sound can be either the spoken word (adding extra story information), sound effects or music. In Figure 4 a page from *The Tortoise and the Hare* is reproduced. After the text has been read, the reader has the opportunity of clicking on any of the animals, which then produces an animated sound sequence; for example, the trio of birds produces a short close harmony trio, the beaver produces a short 'rap' sequence, the frog plays bass guitar on the flamingo's legs, etc.; when the page is turned, all the instruments perform together.

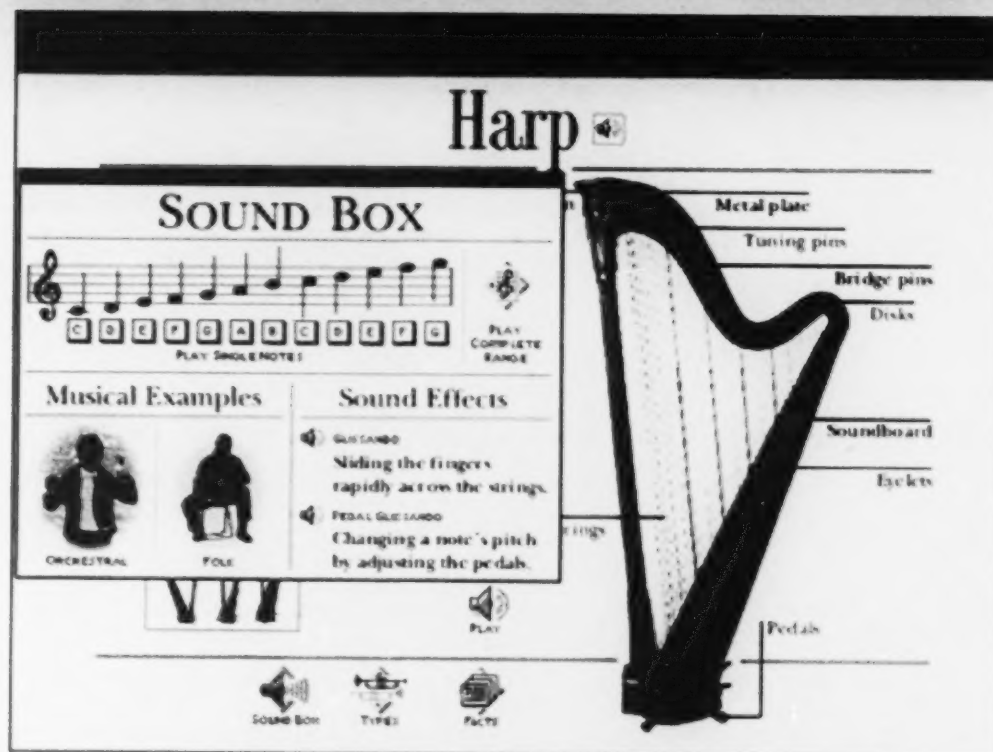


Figure 1: A musical instrument with sound box activated from *Microsoft Musical Instruments*.

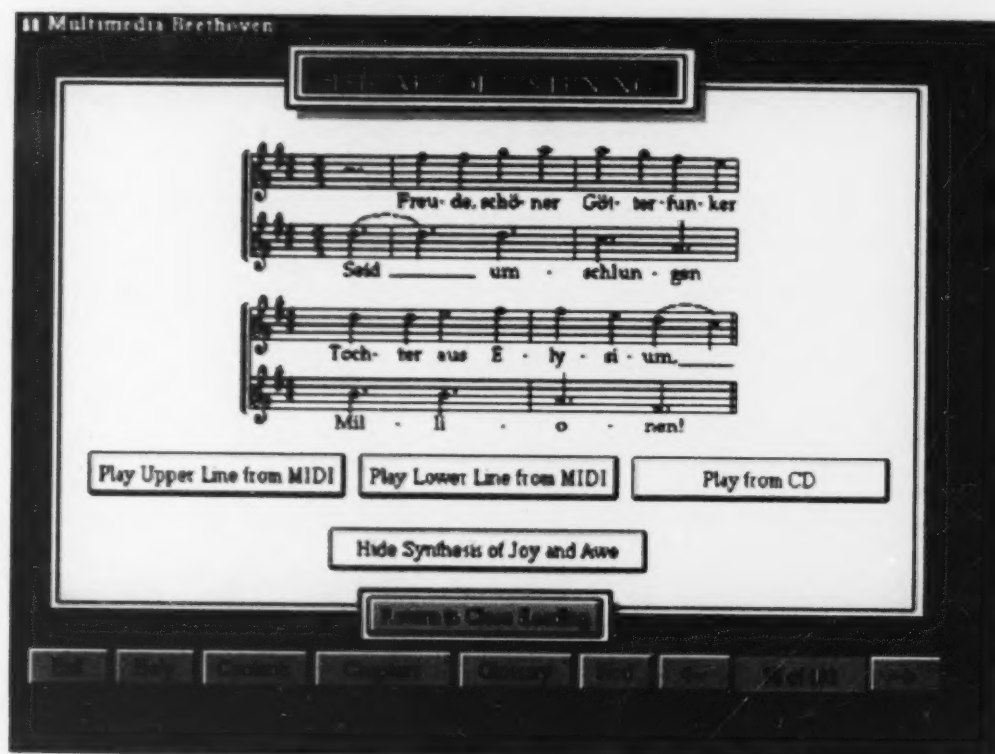


Figure 2: A section of the score from *Microsoft Multimedia Beethoven*.

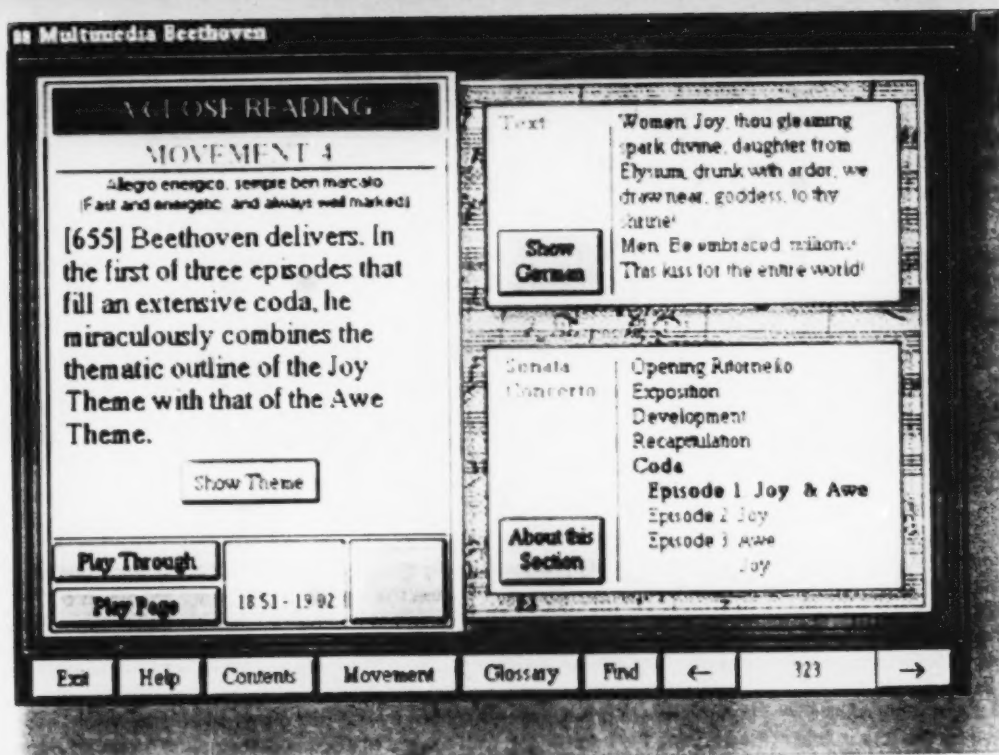


Figure 3: A close reading from *Microsoft Multimedia Beethoven*.

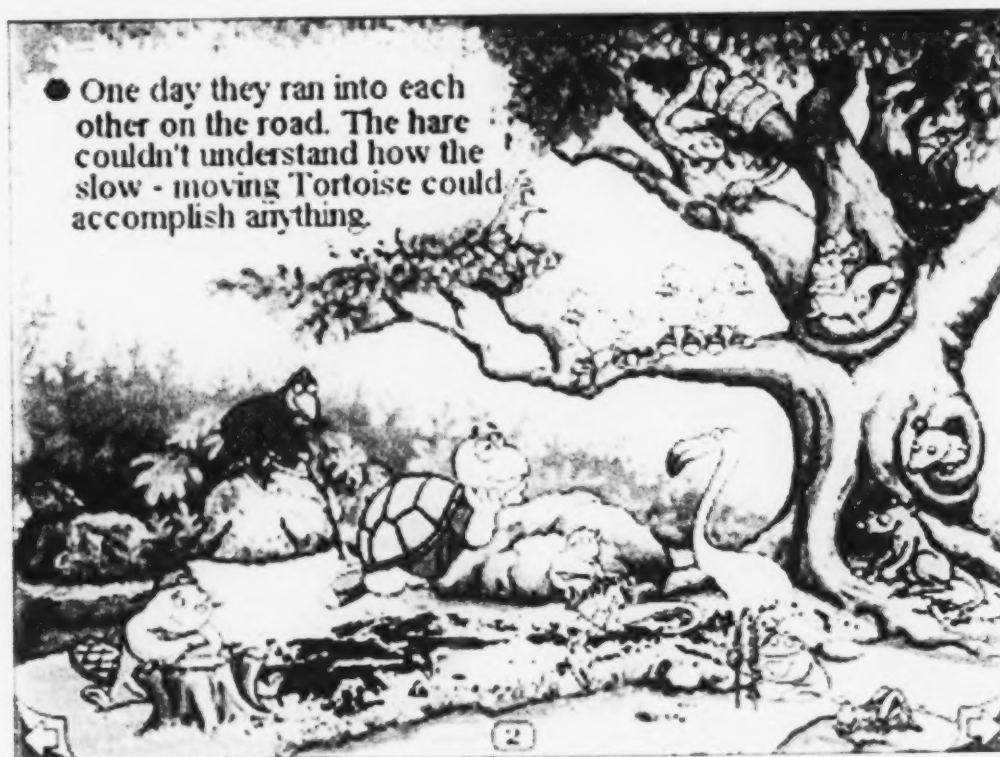


Figure 4: A page from *The Tortoise and the Hare* (Living Books series, published by Broderbund).

One can only expect that more and more multimedia will be added to electronic publications in future, especially when larger storage media become freely available, such as Digital Video Disks (DVD). On the Internet it is becoming easy, and indeed very common, to add animation sequences to documents by means of Java. Since Java programming is platform independent and fairly easy to do, many pages now have animation added to them — very often only gimmicks indicating that the creator of the page has the know-how to do Java programming and the animation adds no value to the publication. The reader should, once again, evaluate the usefulness of the multimedia that has been added by making sure that value has in fact been added to the publication, and that multimedia has not merely been incorporated because the technology was available to do so.

3. Value-adding through the inclusion of tests, games and guided tours

Tests and games can be added at any point in an electronic publication, for example in any section of the publication or at the end. Games may have a purely recreational function but should be structured in such a way that they keep the reader's attention, as well as test his or her knowledge of the subject. Testing can take the form of multiple choice questions, form fill-in, typing single words or short sentences, etc. Proper feedback and references to the place in the text where the problem is discussed should be given, as is done in, for example, the Mindmaze in *Microsoft Encarta* (1996). Tests should comply with the standards for testing in computer based education, which is a separate, complex subject with a detailed literature. This will not be discussed any further; cf., for example, Alessi & Trollip (1991) and Dean & Whitlock (1992) (Ref 3) for details.

Guided tours provide readers with an easy option of navigating through an electronic publication. Instead of having to make choices of which links to follow, the tour guide takes the reader on a predefined tour through the publication. This is obviously against the principle of the reader being in control of the information space, since the only control the reader has is the ability to interrupt the tour at any given point. In a number of the Microsoft Home publications, e.g. *Microsoft Ancient Lands* and *Microsoft Dinosaurs*, the tour guides also provide spoken comments on the textual information, thereby providing additional information not available in the text version of the publication. Very often these comments are purely anecdotal, but sometimes essential additional information is provided by the tour guide; not accessing the information with the aid of the tour guide would therefore deprive a reader of these comments.

Guided tours are sometimes also used to give an overview of how a publication works. By means of such a tour the reader can find out how to make use of all the possibilities in the publication. If this type of tour is essential to enable a reader to find information in the publication, it could be that the structure and functionalities of the publication are not sufficiently transparent.

4. Value-adding through integration of databases

An electronic publication can be read (or queried) as a separate database, but the possibility also exists to integrate various publications by means of hyperlinks between related information in the individual publications.

A well-known example is *Microsoft Bookshelf 1996–1997*, in which eight publications are integrated, viz. an encyclopaedia, atlas, world almanac, dictionary, thesaurus, two volumes of quotations and a book on the Internet. Each of these publications can be read independently, but links between the various publications provide a much richer information environment than the eight publications taken individually. It is for instance possible to look at a map and with the click of a button access information in the world fact book, and from there link to the encyclopaedia, etc. The rich web of hyperlinks within each single publication and between the various publications allow the reader to access the detail and variety of information required for a specific task. The Internet book in *Microsoft Bookshelf* provides the reader with access to information on the Internet, and numerous URLs are provided by means of which the reader may obtain additional information on practically any aspect via the World Wide Web, newsgroups, etc. Updated information can also be downloaded from Microsoft's Web server. *Microsoft Bookshelf* therefore integrates these eight publications with the practically unlimited information space of the Internet.

Biblical text databases such as *BibleWindows* and *BibleWorks for Windows* contain Hebrew, Greek and Latin texts of the Bible as well as various translations; some also contain morphological databases of Hebrew and Greek words, that is databases in which all words in the Hebrew and Greek texts are parsed. The morphological databases can be queried independently (usually by means of a set of menus and forms that can be filled in to simplify the construction of a search significantly). In addition to this the morphological databases are linked to the Hebrew and Greek text databases: by selecting a word in the Hebrew or Greek text, the morphological analysis is displayed. Double-clicking on a word in *BibleWorks for Windows* opens a window which displays the morphological analysis of the word, as well as an optional dictionary entry; cf. Figure 5 for an example.

Links to semantic databases — simple word lists or state-of-the-art lexica — are also possible. For instance, *BibleWindows* contains links to an electronic version of the Greek-English lexicon of Louw & Nida (1988) (Ref 4). Each publication, viz. the text databases and the lexicon, can be accessed separately. The dictionary can therefore be used as a standalone dictionary for standard use. However, selecting a word in the Greek text database allows the possibility of looking up the relevant word automatically in Louw & Nida; cf. Figure 6 for an example of *BibleWindows* with a link to the Louw & Nida lexicon. *BibleWindows* uses its own (proprietary) database formats and

search engines in the text databases, but, through morphological analyses of texts and hyperlinking, access to the Louw & Nida lexicon which is programmed in FolloViews is automatic. For more details see Bothma 1994 (Ref 5).



Figure 5: Hebrew and Greek morphological analyses in BibleWorks for Windows (published by Hermeneutika).

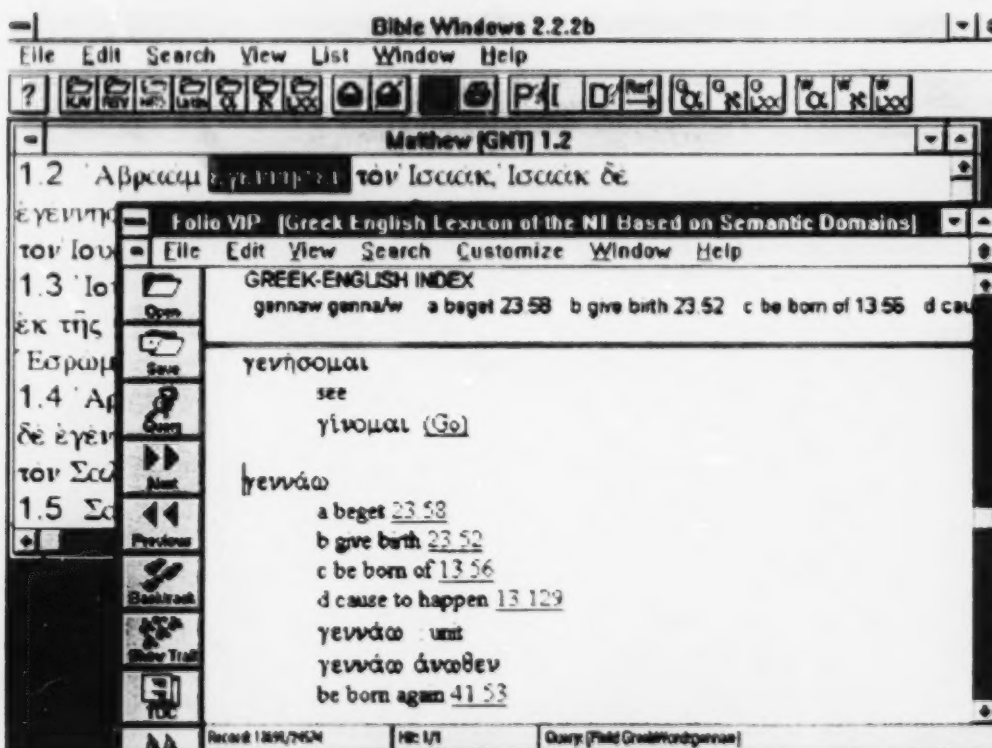


Figure 6: Linking BibleWindows to the Louw & Nida lexicon (published by Silver Mountain).

5. Value-adding through improved structuring

5.1. Hierarchical structuring of information

Electronic publications can be structured hierarchically in terms of the macro level and/or the micro level.

At the macro level this implies that, at the top level, the reader is usually supplied with a number of choices; selecting any of these would take the reader one level lower in the hierarchy, where a series of further selections is presented, until the required information is displayed, be this an article in an encyclopaedia or a specific chapter or paragraph in a book. This is commonly used in the Microsoft Home series (for example *Ancient Lands*, *Dangerous Creatures*, *Musical Instruments*, *Dinosaurs and Ocean*) and the Dorling Kindersley multimedia publications (such as *The Way Things Work*, *Encyclopedia of Science*, *Stowaway!*, and *The Ultimate Human Body*).

If an entry in especially a dictionary or encyclopaedia is long, a microlevel hierarchical structure is often also supplied. This means that the article can be collapsed to an outline of the headings and subheadings of the various paragraphs in the article, or that these headings can be extracted to a separate window. By selecting any of the headings, the full text of the paragraph is displayed. These techniques are employed very effectively in *Microsoft Encarta*, where proper paragraph headings clearly indicate the contents of the paragraph, and *Microsoft Cinemania*. In the *Oxford English Dictionary* this technique is unfortunately employed very ineffectively, because of the lack of proper headings.

A combination of macro and micro level hierarchical structuring occurs when the reader is presented with an electronic table of contents which can be expanded or contracted; this table of contents is then linked directly to the full-text. The screen is divided into a table of contents section and a full-text section; by clicking on any item in the table of contents the reader directly moves to the relevant passage in the full-text; cf. Figure 7 for an example taken from the *Text Encoding Initiative Guidelines P3*.

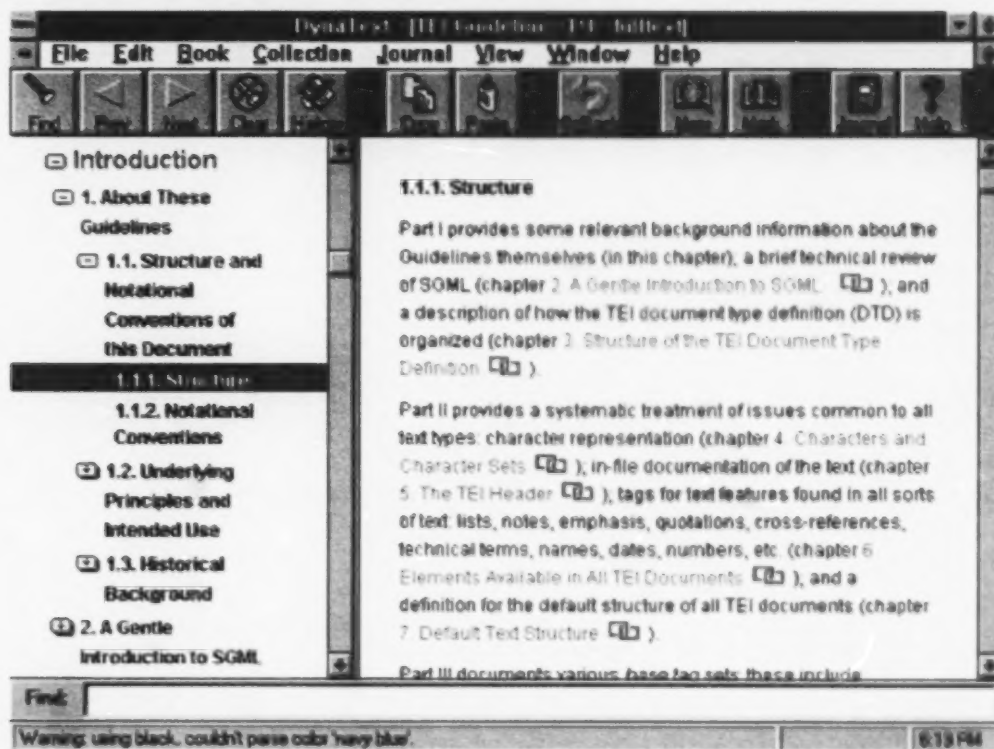


Figure 7: Electronic table of contents linked to the content in the TEI Guidelines P3 (electronic version published by Electronic Book Technologies).

5.2. Hyperlinking to related/associated material

Many electronic publications support hyperlinking between a word/phrase or paragraph and related materials. This can be implemented in two ways, viz. by explicit referral and by embedded links in the text.

Users can be referred to related material explicitly by means of a reference list listing related material at the end of an article, or a 'See also/Related articles' button.

Embedding links in the text to related material implies that the reader may at any stage acquire related or explanatory material on demand. This means that the reader can choose which trails of information he or she

would like to follow and the information can be accessed in a non-linear, non-sequential manner (Ref 6).

Embedded links are usually indicated by means of some or other typographical convention. The text of the embedded link may be displayed in another colour, in inverse video, in a specific typeface or underlined, or a combination of these features. Most Microsoft publications use a second colour together with underlining; full underlining indicates a cross reference to another article, whereas broken underlining indicates glossary information available in a pop-up window.

Embedded links need not necessarily be indicated, and especially in dictionaries all terms can act as the anchor point of an embedded link. By selecting a term in a dictionary the reader is then taken automatically to the lemma of the selected term. In the *Oxford English Dictionary* this is done by selecting a term and then activating the button 'Xref', and in Macmillan's *Dictionary for Children* clicking with the right mouse button on any word takes the reader to the lemma of that word.

5.3. Navigational aids

Readers can browse (or navigate) through the information by means of the above-mentioned hierarchical structure or by following links. Various navigational aids (Ref 7) to help the reader maintain an idea of where he or she is in the electronic publication are usually also available, such as:

- a backtrack facility by means of which one can retrace the path one has followed through the information;
- a 'home' button which enables the reader to go to the opening screen of the publication in one step;
- a history facility which can display a list of all the sections that have been accessed; and
- electronic bookmarks which are user-defined.

Absence of such devices can easily lead to the user getting lost in hyperspace.

Hierarchical structuring of information, hyperlinking to related information and navigational aids enable readers of electronic publications to find relevant information (the correct information at the detail required) much easily and much more efficiently than in paper-based equivalents. The possibility to do full-text searching further enhances an electronic publication.

6. Value-adding through full-text search facilities

Using an index is the main method of locating information in a printed book (in addition to the table of contents). It is usually constructed on the basis of predefined terms and contains only selective references to the text. It is therefore limited in terms of completeness, both with regard to the list of search terms as well as with regard to the exhaustiveness of a specific search. Adding full-text search facilities in an electronic publication can vastly improve access to information. Full-text searching implies that the full text of the publication is searched for the specified term(s), resulting in a comprehensive list of occurrences of the term(s).

Searches for a single word can be done, as well as for phrases. Boolean searches are supported, i.e. words and phrases can be combined with the logical operators AND, OR and NOT. The use of wildcards and truncation is usually also possible, as well as the use of proximity operators.

An electronic publication supports searches not possible in the equivalent paper-based publication. Paper-based dictionaries and concordances allow users the possibility of single word look-up, according to the alphabetical structuring of the publication. An electronic dictionary such as the *Oxford English Dictionary*, however, allows searches on the etymology of words, quotation searches, author and title indexes, etc. in addition to full-text searching; cf. Figure 8 for an example.

How full-text search features are implemented differs from publication to publication. In some cases a command line interface is employed, in others a menu-driven search construction. Some programs even allow the user to select the search method best suited to the specific user, for example in the *English Poetry Database* published by Chadwick-Healey.

Command line interfaces can be fairly complex to the average end-user. One example will suffice. In *BibleWorks for Windows* — a concordance program containing various translations of the Bible as well as the texts in Hebrew, Greek and Latin — users can do complex Boolean searches combined with truncation and proximity operators. To search for any derivative of 'faith' occurring within two verses of any derivative of 'work', and this combination occurring within two verses of 'Jesus' or 'Christ', one has to enter the following command at the command line (Ref 8): `(.faith* work*;2).2(/jesus christ).`

This is too complex for the average user since not only does one have to understand the principles of Boolean searching, but one would also have had to learn a difficult command line interface (which differs from program to program) in order to be able to construct a valid search command. Consequently most publications tend to offer various types of help menus to construct full-text searches.

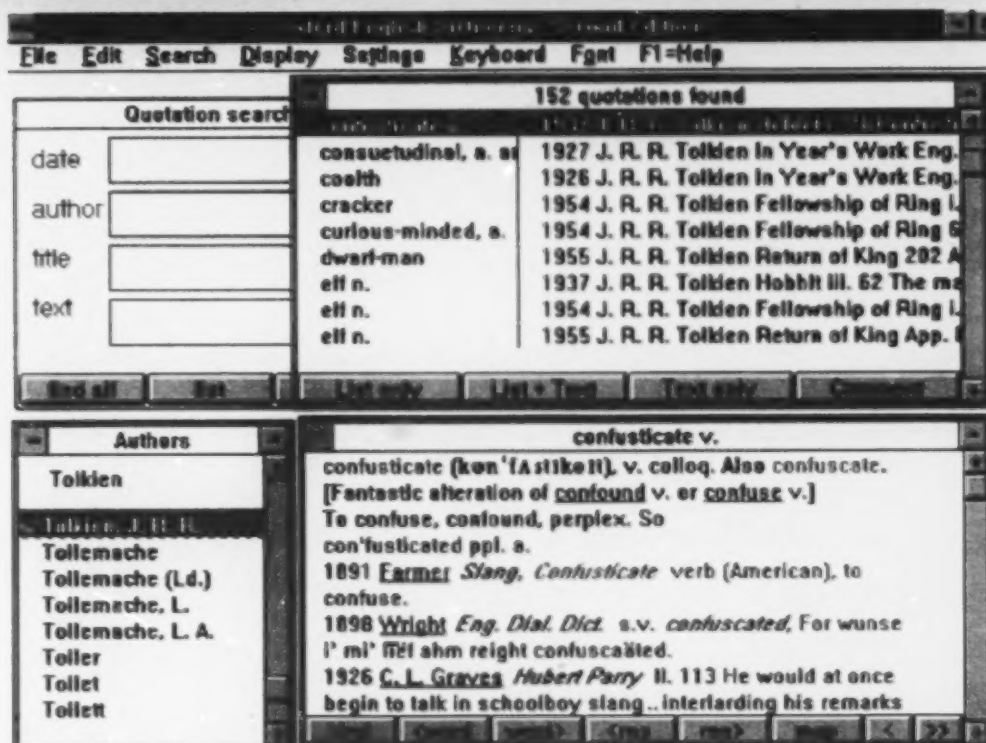


Figure 8: Quotation search on authors in the Oxford English Dictionary.

6.1. Help facilities to construct full-text searches

The full-text search facilities discussed in the preceding section make access to information in electronic publications much more sophisticated than is possible in equivalent paper publications. However, to construct Boolean searches with proximity operators and so forth can be fairly complex and publishers should provide help screens to simplify the process. Again there is no standardisation in the help interface of different publications, and the menus sometimes vary from a simple list that provides the option to specify a number of search terms and select AND or OR as operators, to highly sophisticated menus which can do searches across various fields or search wizards, by means of which the search can be built up by answering a number of questions.

Figure 9 provides an example from the *Andromeda Interactive Space Encyclopedia* which offers a fill-in list to specify the search terms and a pull-down menu to specify AND or OR, as well as the possibility to specify proximity in terms of the word range within which the specified terms must occur.

In *Microsoft Encarta* various other help facilities are available to construct a search. In addition to the primary search term(s) (combined with Boolean operators), it is for example also possible to define the broad category which has to be searched, the date range (or precise date), the locality, etc.; these features are available by clicking on the various buttons on the search screen, which then activate a sub-menu; cf. Figure 10 for an example. The program could also assist the reader in filling in the forms by means of a search wizard. *Microsoft Cinemania 96* offers similar features by means of which a very complex search strategy can be formulated.

It is quite evident that full-text search facilities provide much more sophisticated access to information than is possible by means of an index and a table of contents in a paper-based publication.

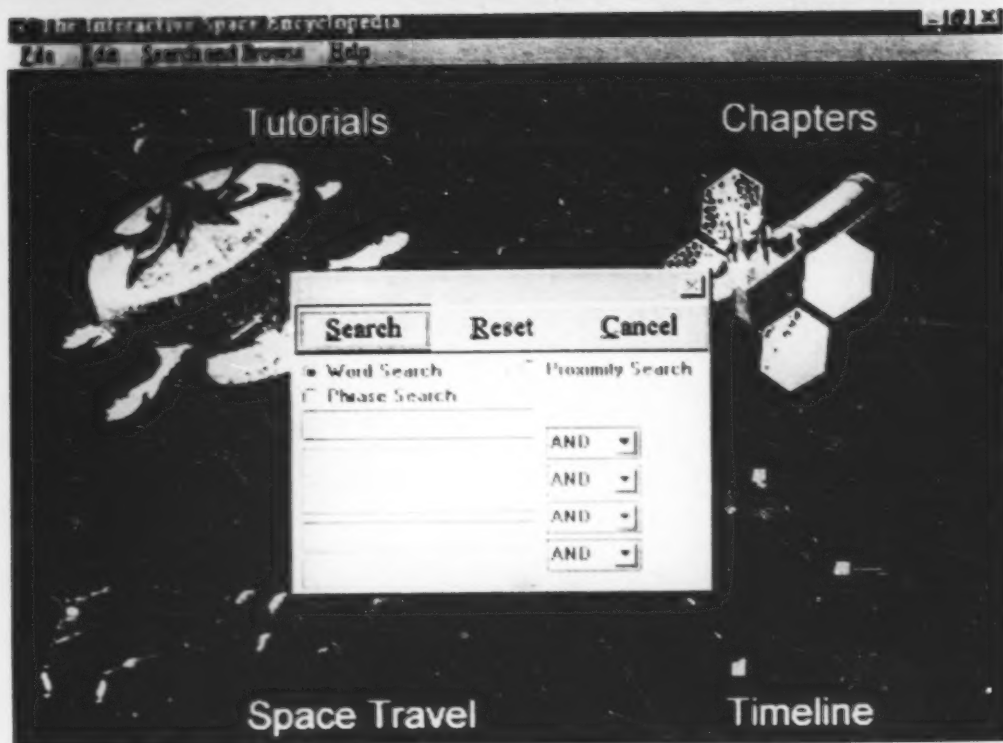


Figure 9: Help to create Boolean searches in the *Interactive Space Encyclopedia* (published by Andromeda Interactive).

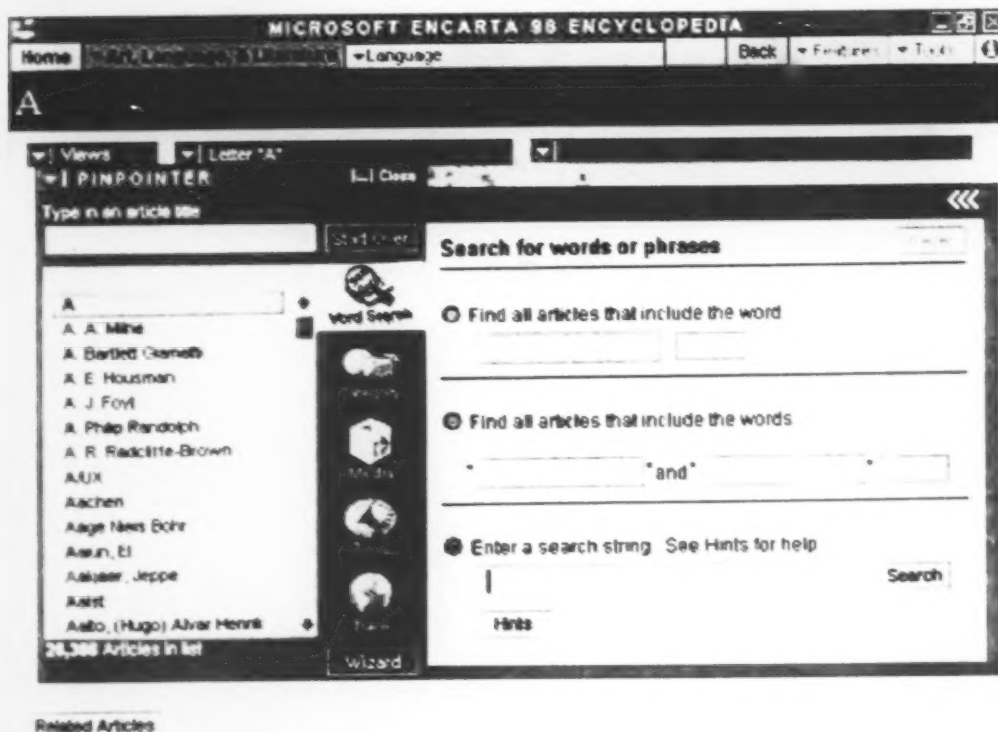


Figure 10: A screen showing some of the search facilities in *Microsoft Encarta 1996*.

7. Conclusion

Obviously not all methods of adding value to electronic publications are relevant in all cases. It is for instance not logical to add sophisticated search facilities to a story book for children; similarly, the addition of sound and animation may not be relevant in an electronic publication of the classics of English literature.

It is, however, important that all readers — end-users, information scientists and librarians alike — should know how value can be added in electronic publications, to enable them to judge whether any features that have been added to an electronic publication make the publication more useful than an equivalent paper publication, or whether these features are only gimmicks which have been added because the technology is available.

The perception very often exists that electronic information sources are of a higher quality than paper-based information sources. This is not necessarily true. It has to be stressed that every electronic publication should be evaluated on various levels, viz. the level of content, as well as technological aspects such as structural properties, the various methods provided to gain access to the information, and visual aspects. The best interface, design and value-adding technologies cannot make a publication useful if the content is not sound, i.e. correct, complete, up-to-date, relevant, etc.: technology cannot improve bad content. On the other hand, if the content is good but technology has not been implemented in the optimum way, the publication would suffer. This obviously brings a whole new dimension to the work of anyone who has to design or evaluate publications.

All readers should be aware that many excellent electronic publications are available, but that there are also many titles which are but poor imitations of quality electronic publishing.

Prof. Theo J D Bothma
Department of Information Science
University of Pretoria
Pretoria 0002
South Africa
Tel.: +27 (0)12 420 2293
Fax: +27 (0)12 342 2012
E-mail: bothma@libarts.up.ac.za

References

- [1] Commission of the European Communities (1993) *New Opportunities for Publishers in the Information Services Market*, Main report, Publication no 14926 EN, Luxembourg, p. C7.
- [2] Feldman, T. (1991) *Multimedia in the 1990s*, British National Bibliography Research Fund, Report 54, The British Library, pp. 105, 107.
- [3] Alessi, S.M. and S.R. Trollip (1991) *Computer-Based Instruction: Methods and Development*, 2nd edition, Prentice Hall, Englewood Cliffs, NJ; Dean, C. and Q. Whitlock (1992) *A Handbook of Computer-Based Training*, 3rd edition, Kegan Page, London.
- [4] Louw, J.P. and E.A. Nida (1988) *Greek-English Lexicon of the New Testament Based on Semantic Domains*, United Bible Societies, New York. Electronic version published by Silver Mountain Software, 1994.
- [5] Bothma, T.J.D. (1994) The use of information technology in studying biblical languages, *Journal for Semitics*, 6(2), 201-230.
- [6] Hypertext and hypermedia are very powerful methods by means of which information can be organised. For more details cf., for example, Nielsen, J. (1995) *Multimedia and Hypertext. The Internet and Beyond*, AP Professional, Boston; McKnight, C., A. Dillon and J. Richardson (1992) Hypermedia, *Encyclopedia of Library and Information Science*, vol. 50, Supplement 13, Marcel Dekker, New York, pp. 226-255; McKnight, C., A. Dillon and J. Richardson (1991) *Hypertext in Context*, The Cambridge Series on Electronic Publishing, Cambridge University Press, Cambridge.
- [7] Providing sophisticated navigational aids are standard organisational devices in hypertext; for additional navigational aids cf., for example, Nielsen, J. (1990) Navigating through hypertext, *Communications of the ACM*, 33 (3), 297-309 and the items cited in Ref 6 above.
- [8] There is unfortunately also a disadvantage to this type of concordance program: no morphological analysis or tagging of the English is done and the user searching for 'go' and its derivatives, such as 'goes' or 'going', will have to use a wildcard (e.g. 'go*'), but will then also find any other word beginning with 'go' such as 'God', 'gods', 'gospel', etc., while not locating occurrences of 'went', which will have to be entered separately.



Resource management in the electronic environment: company library Web pages and collection development principles

Hazel Hall and Aileen Russell

Queen Margaret College Edinburgh, UK

Abstract: *Traditional information services provision relies on the planned and logical building of a collection to meet current and future user information needs. This is achieved through the acquisition and discarding of material, which might ideally be based on a regularly reviewed collection development policy statement. The evolution of collection development policies facilitates discussion between user groups and service providers, bringing information professionals closer to those whose needs should be served, and assures that the materials selected and serviced by the information unit are worth the investment made in them.*

Information professionals whose careers have been based in library services are familiar with the mechanics of collection development and their application. Those now involved in the creation of company library Web pages, for providing user access to remote Internet resources as a complement to the physical collection held locally, may decide to devise collection development policies specific to Web page management. The priorities of a Web page collection development policy, and of the people involved in its planning and implementation, differ from those of a traditional (print) collection development policy. This raises issues of concern with regards to serving user information needs now and in the future.

Keywords: collection development, World Wide Web, Internet, home pages

1. Introduction

This paper discusses the principles of collection development policies for print and electronic information resources as described in the library and information management literature. The results of a telephone survey on the degree to which British company libraries are networked, and a subsequent questionnaire on the development of collection development policies completed by those offering library Web pages, are presented.

2. Collection development and collection development policies — definitions

2.1. Collection development

Collection development has been described as 'a planning function for the systematic and rational building of a collection' which 'relates to the selection and acquisition of material for an expanding collection, and decisions on the material to be included in that collection' (Ref 1). As well as considering the addition of new materials, the maintenance of a live, balanced, up-to-date collection depends upon the withdrawal of items which are used infrequently or are obsolete (Ref 2).

2.2. Collection development policy

The collection development policy is a stated description of 'the collection as it is now, as it will be developed, and defines the rules directing that development' (Ref 3). As a working document it is reviewed and revised continuously.

3. Policies with a purpose

There are two main reasons for devising and running a collection development policy.

3.1. Institutional goals, changing service requirements and meeting user information needs

A priority of the librarian is to ensure that the information needs of the user community can be met. In the company library context, the information professional contributes to the organisational effort devoted to achieving business objectives. Although this can be achieved without formalised collection development activity, following a documented policy can demonstrate more clearly how to address user information needs now and in the future, aligning service provision more closely with organisational goals (Ref 4).

3.2. Financial accountability

As financial accountability becomes more important (to check sensible spending, as well as ward off the threat of closure or out-sourcing of information services), planned resourcing of an information service is crucial. This 'informs a library's governing and funding body about the library's directions and provides a clear and carefully articulated rationale for its collection' (Ref 3).

Snow (Ref 5) notes how 'Along with a commitment to service and resistance to censorship, the necessity of the written collection development policy is part of the creed of librarianship' and demonstrates that 'as with many creeds, this one is honoured as much in the breach as the observance.' His review of the literature uncovered only two articles critical of the practice, and a large number which warranted the adoption of collection development policies for no other reason than the fact that they were 'essential'. This kind of unjustified statement can be found in the discussion of the electronic environment: 'Collection policies and procedures must be created for local developers of virtual libraries and for co-operative virtual collection programs just as they are for physical library collections' (Ref 6).

4. Policy parts

The components of a collection development policy are outlined fairly extensively in the library and information management literature (for example, Ref 2). General objectives outline the introduction and scope of a collection development policy (Ref 7). More specific issues addressed include:

- identification of the clientele to be served;
- subject boundaries of the collection;
- priorities and limitations governing acquisitions (e.g. selection, duplication, replacement, orders, donations and exchanges);
- cataloguing;
- retrospective conversion;
- preservation;
- fund-raising and budgeting;
- discard procedures (focused on the removal of older, outdated and inaccurate information sources from the collection, taking into consideration the age of the material, its physical condition, space restraints and availability of better sources);
- provision for periodic review and revision of policy;
- statements of present practice and future intentions.

5. Format-specific collection development issues

5.1. Special formats

The basic principles of a collection development policy should apply to *all* media managed by the library (Ref 8). However, extra provision needs to be made for sources which require special treatment due, for example, to size, durability, handling or means of access to the information. Newspapers and microforms fall into this category (Ref 7). For collections that include electronic media the question of equipment and staffing assume greater importance. The policy covers criteria relating to the information source *per se*, as well as to the devices which support its function with relation to cost, user interfaces and physical space, etc. (Ref 4). The collection developer also discovers a closer relationship with technical staff (Ref 9). It is not easy to find discussion of the creation and implementation of collection development policies for 'non-traditional' media. On completion of a survey of 69 academic libraries in the United States, Boyarski & Hickey stated that 'electronic formats and services, current

hot topics in librarianship, are seldom mentioned.' In the sample policies provided in their text, allusions were found to locally produced media software, media equipment, instructional resource equipment, database services, CD ROM and networking for inter-library loan purposes (Ref 10).

5.2. General Web page design guidelines

Numerous sources, from downloadable files on the Internet to articles in magazines sold in corner shops, provide advice on Web page development. However, the focus of these is on the construction of page layout with regard to the use of HTML, graphics, interface design, link labelling and so on, rather than collection development issues. Drawing an analogy with a physical library, it is as though the presentation of the stock (through shelf arrangement, guiding and classification) is deemed more important than the information sources actually held.

5.3. Web page design and collection development literature

In the library and information management press, articles on Web pages come closer to considering collection development. For example, Falcigno & Green (Ref 11) provide practical hints in their paper on setting up a set of Web pages for a company library. These include establishing user information needs, adding new links as appropriate, checking the validity of existing links, allocating staff responsibilities and taking into account the legalities of providing information on the network. They counsel against adding links to 'interesting' sites: each resource identified for the Web page must fit in with the library's collection development policies. At the Untangling the Web conference held in Santa Barbara in Spring 1996 several papers related to Web page development and maintenance were presented (Ref 12). Tennant's paper on Web server management discusses user needs, fostering data owners and the formulation of Web policies. It also points to other resources of relevance to policy design and implementation (Ref 13). A number of academic libraries have provided accounts of managing their Web pages (Refs 14-16). An extensive listing of higher education and research library Web sites in the UK is included in the Exeter library and related resources guide (<http://www.ex.ac.uk/~itilsed/lib/uklibs.html>).

6. British company libraries and Internet resources

For the purposes of the research project described in this paper a number of company librarians were approached. A systematic sampling method was employed to identify respondents from a directory (Ref 17) and subsequently a telephone survey was conducted with 56 individuals. The initial interviews sought to establish the degree to which British company librarians have access to Internet resources, and to seek subjects for a follow up questionnaire on library Web page development.

It was found that 73% (41) of the company libraries had Internet connections. The Internet was regarded as a useful tool for communication (e-mail), for answering reference enquiries and sourcing grey literature such as product information. Of the networked libraries just four (7%) offered internal library home pages. However, 15 (i.e. just over a quarter of the total sample) planned to set up a home page in the near future. A number of librarians interviewed readily expressed misgivings about using the Internet as part of their information services portfolio, making reference to its lack of structure and control, fears about security and over-reliance on a system that might go down. This kind of comment has been typified in other recent research (Ref 18).

7. Company library Web pages and collection development policies

The four respondents who admitted that their library maintained a set of home pages agreed to complete a follow-up questionnaire. Of the four, three operate a collection development policy for their Web pages: one as part of the main policy, another as an independent policy and the third as a special library project. Two libraries also have stated policies for non-Web resources.

It was felt that adhering to a set of collection development guidelines is more important for networked resources than for the local physical stock. There were several reasons why the librarians advocated this, the most important being to provide clear accountability to finance managers. Running a Web-based service has visible implementation and maintenance costs for software and hardware. Spending has to be seen to be directed sensibly. The notion that 'if we can get it on the Internet it's free' is meaningless to those signing the cheques for new terminals or watching the daily Internet activities of information services staff.

At a more practical level, the aim of providing an organised collection of working links for users was identified as the second main reason behind referring to a documented collection development policy. The librarians interviewed were conscious of the wastefulness of browsing for information on the Net, and sought to use their skills in establishing some form of bibliographic control on the resources known to be useful to their user group. In all three Web collection development policies the criteria for maintaining consistency of style of the pages were also stated.

In the librarianship literature the most regularly quoted purpose of maintaining a collection development policy is to ensure that user information needs are met. Secondary to this is the notion that libraries can use their policy to ensure sensible spending. It would appear from these results that the relative importance of these two

arguments is switched with relation to Web page collection development. Perhaps this is because offering a Web-based service is regarded as a luxury and, as such, has to be justified in financial terms before user information needs can be considered. As far as print sources are concerned, financial support for their management is more readily assumed; therefore focus falls on user information needs first.

A further difference between the operation of Web and other collection development policies is the number and type of people involved. In the sample firms only library staff and users had input to the collection development policies for print sources. For electronic sources (other than those accessed via the Web pages) administrators and staff became involved. With regard to Web page collection development policies, the library user group's involvement was replaced by that of managers and the organisation's PR function. It could be argued that when the media of the information moves from the traditional to the new (print to microform to CD-ROM to Web), more people join in the discussion of its management. As cost and technical dependence increase, so does the interest of management and technical staff. This might be to lower the risks of procurement mistakes, to meet with company information and PR strategies, or simply because library staff do not have confidence in their technical abilities. That the users had no involvement in the design of collection development policies for the Web pages in the sample here may be a cause for concern. How can a library service meet user information needs if they are not known and documented?

8. Rejecting the creed: why collection development policies are not devised

A straight reading of the librarianship literature would give the impression that libraries without collection development policies would be rare. However, all the case study firms identified at least one medium for which there was no collection development policy, demonstrating that 'its [collection development policy] acceptance in practice, as opposed to in theory, is by no means universal' (Ref 5). One respondent explained that there is no point in devising elaborate formal documents to help maintain a small subject-specific collection. Taking into account the number of individuals involved in the non-print policies described above, it might be the case that the effort of drawing up such a policy is simply not justified (Ref 19). It could be argued that this leads to unplanned, reactive working practices. However, librarians in this sample explained that user recommendations, customer demand as well as source availability, led their collection development for a media where a policy was not in operation. An alternative approach to the formal document replaces policies condemned as 'static, reactive and of little practical utility' that 'have outlived their purpose' (Ref 20). This might be preferable to upholding a policy which does not 'belong' to the users whose needs are to be served, as is the case in the three Web page policies of the sample firms. More serious are the dangers where budgets are erratic and service provision depends on the future upgrading equipment. An agreed policy alerts financial managers of future needs.

9. The future of collection development in the electronic environment

That electronic sources and their management are becoming more important in the company library context was evident in the words of one respondent who outlined selection policy: 'if you want to buy it, ask about electronic versions. If it is not too expensive or complex, we'll buy that in preference to print.' One respondent explained that the aspects of collection development activity (outlined in Section 4 above) will continue to be important, with a greater need for library staff to have a better awareness of technology to secure 'wider, more technically relevant coverage of the technically dense material required'. The format of the information sources managed does not change the general tenet of the librarian: 'we fit in the same way as we always have — by providing evaluation and collection expertise' (Ref 11). This should result in a balanced collection of physical stock, alongside 'owned' and remote electronic sources.

10. Conclusions

This study parallels others (for example Ref 18) in demonstrating that the Internet is yet to be adopted wholeheartedly by the British company library community. The disorganised and uncontrolled nature of the sources discourages its use as a valid resource. It might be regarded as ironic that those who *do* have the skills in evaluating and organising information sources are not yet at the stage of implementing plans to provide a useful interface to the Internet for their users through a set of internal library Web pages. Here lies an opportunity to promote the library, demonstrate its commitment to information provided in all formats, and deliver further networked services over both the intra- and Internet. This is of particular value to companies that operate across countries and time zones. Perhaps the reluctance is due to the perceived enormity of the tasks of implementation and maintenance of Web pages: the identification of material, its evaluation, organisation and so on, as well the necessity to seek training and train others. A fear of creating excessive demand may

also be stalling development. Witness what the introduction of CD ROM abstracting and indexing services did to inter-library loan requests.

Although the initial purpose of this research project was to assess the applicability of traditional collection development principles to Web page development, concerns regarding the late involvement of company library staff in setting up internal home pages assume importance in the conclusions. The results of the questionnaire demonstrate how the user has dropped out of the planning of collection development policies when it comes to Web pages. Might we see a situation where library staff involvement is relegated in favour of that of systems or marketing personnel? The opportunity for library staff to share their expertise in the basic principles of library service, collection development and preservation with other professionals should not be missed.

Hazel Hall
Department of Communication and Information Studies
Queen Margaret College
Edinburgh EH12 8TS
UK
Tel: +44 (0)131 317 3509
Fax: +44 (0)131 316 4165
E-mail: cihall@cis.qmced.ac.uk

References

- [1] Line, M. et al. (1994) *Librarianship and Information Work Worldwide*, Bowker-Saur, East Grinstead.
- [2] Gates, J.K. (1993) *Introduction to Librarianship*, Neal-Schuman, New York.
- [3] Johnson, P. (1994) Collection development policies: a cunning plan, *Technicalities*, **14**(6), 3-6.
- [4] Jenkins, C. and M. Morley (1991) *Collection Management in Academic Libraries*, Gower, Aldershot.
- [5] Snow, R. (1996) Wasted words: the written collection development policy and the academic library, *Journal of Academic Librarianship*, **22**(3), 191-194.
- [6] Baldwin, C. and S. Mitchell (1996) Collection issues and overview. Available at <http://www.library.ucsb.edu/untangle>.
- [7] Magrill, R.M. and D.J. Hickey (1976) *Acquisitions Management and Collection Development in Libraries*, American Library Association, Chicago.
- [8] Demas, S. (1994) Collection development for the electronic library: a conceptual and organizational model, *Library Hi-Tech*, **12**(3), 71-80.
- [9] Buckland, M. (1994) What will collection developers do? *Information Technology and Libraries*, **14**(3), 155-159.
- [10] Boyarski, J. and K. Hickey (Eds.) (1994) *Collection Management in the Electronic Age*, American Library Association, Chicago.
- [11] Falcigno, K. and T. Green (1995) Home page, sweet home page: creating a Web presence, *Database*, April/May, 20-28.
- [12] <http://www.library.ucsb.edu/untangle/>
- [13] Tennant, R. (1996) The art and science of Web server management. Available at <http://www.library.ucsb.edu/untangle>.
- [14] Furner, J. (1995) The World Wide Web in libraries: an overview, *Vine*, **99**, 3-18.
- [15] Flowers, K.A. et al (1995) Collection development and acquisitions in a changing university environment, *Library Acquisitions: Practice and Theory*, **19**(4) 463-469.
- [16] SUL/AIR (March 22nd 1996) *Web Advisory Committee—Charge and Membership*. <http://SUC3.stanford.edu:8080/wac/wacchrg.html>.
- [17] Dale, P. (1993) *Guide to Libraries in Key UK Companies*, British Library, London.
- [18] Kelly, S. and D. Nicholas (1996) Is the business cybrarian a reality? Internet use in business libraries, *Aslib Proceedings*, **48**(5), 136-144.
- [19] La Guardia, C. (1992) Electronic databases: will old collection development policies still work? *Online*, **16**(4) 60-63.
- [20] Hazen, D.C. (1995) Collection development policies in the information age, *College and Research Libraries*, **56**(1), 29-31.



The Internet: the demise of traditional online

Heinz Ochsner and K. Thomas

Knight-Ridder Information, Switzerland

Abstract: For end-users the Internet has brought about a revolution in the concept of information provision and opened up a diverse new market place. Though very positive, the proliferation of Web sites and search tools now available to the user can leave a feeling of disorientation. Many dilemmas arise, such as: how to measure the quality of data retrieved? how to locate a good source? and how to develop effective searching techniques? For traditional online hosts the Internet is significant as it represents the first standard platform which does not require the user to make costly technical investments. All products and services developed in HTML are accessible by any Internet user. This new technology has facilitated the breakdown of publishing barriers. It provides an instant medium for authors and information providers to deliver data directly to targeted audiences, bypassing the channel that traditional online has provided. The need for — indeed the role of — online services is challenged and it raises the question: is traditional online a dying industry?

Traditional online hosts have had years of experience in developing and refining sophisticated search tools, but most importantly hosts provide users with standardisation, i.e. the means to search efficiently across a range of sources with the knowledge that the information available is authoritative, evaluated and of real value. Working with information providers, to standardise and index data, hosts can provide a 'one-stop shop'. Using HTML the Internet now provides hosts with the first opportunity to deliver that experience to a wider market, in a style which is compatible with their existing working environment. Traditional online is in decline but hosts are far from becoming extinct. Host services on the Internet are removing the need for users to learn complex searching techniques and the need to go to multiple sites to locate information. The Internet is an opportunity for hosts to deliver targeted tools which precisely match the searching requirements of different audiences. New services are helping to eliminate the current dilemmas and hosts play an important role in that progression.

Keywords: online hosts, information providers, internet, HTML, standardisation, quality data

1. Introduction

Conjure up an image of a memorable journey by car: remember the scenery, the smells, the emotions, maybe even the game of 'I spy'. How many of those memories come from a journey on the motorway (autoroute)? The probability is very few. A journey via a country lane may be memorable and enjoyable but it's easy to get lost or stray from the route, so that when you finally reach your destination (if at all) it has all taken far too long. Country lanes quickly become congested and fail to meet the objectives of a journey. Travelling by motorway, on the other hand, may not be memorable but you complete your journey quickly and easily, taking the most direct route available. Country lanes are idyllic but not for those who want *speed* and *reliability*.

This scenario can be compared to the Internet. The world of new sites to browse through provides an attraction similar to a country lane, where users stop to wander and explore. However, as more and more people try to take part in the new Internet revolution, the more congested it becomes. Traditional information users, familiar with online services that provide precise, authoritative results, have been used to experiencing the benefits of motorway driving where the clear objective of getting from A to B is efficiently achieved. Yet the speed and accuracy of the motorway cannot be found on a country lane. Users seek the best of both worlds. This paper aims to explore the pros and cons of the Internet, to discover what role, if any, hosts will play in the new information environment and question to what extent traditional online services are in decline because of the Internet.

Is there evidence for this decline and demise? Yes, although not obvious, it is there:

- Alta Vista, the best Web search engine of all, with intelligent retrieval akin to some DIALOG commands, has done a terrific job of winning over the punters and is now ready for its stock market flotation (Ref 1);
- Free Web access to Medline, once the cornerstone of many host services and CD-ROM vendors, is out there (Ref 2);
- Flat-fee unlimited access to BIOSIS via JANET is new for 1996 (Ref 3).

All these innovations must be nibbling at the edges of the traditional hosts' earnings or carving out new territory. A clear response is called for.

An international provider such as Knight-Ridder Information has ongoing growth in European countries and the Far East, but a mature market in its American backyard. The bread and butter science citation searching of old is giving way to a broader range of business information demands. So growth in one area is masking and compensating for the decline that exists in another. A large product offering and an international customer base provide protection and keep a company buoyant, if not complacent.

2. Benefit of the Internet to end-users

The Internet has ridden a tide of media hype which in its turn has helped to position the Internet as a revolutionary tool and has helped to educate new users. The Internet is significant more for what it represents than what it actually is. The Internet has opened the gates to a new wave of information provision to new markets. The Internet represents a distinctive change in the industry of information provision and communication. It brings together markets which have traditionally never been together. For example the entertainment market and the consumer market have not traditionally been involved in online services and could never be found in the same place as business information but now, whatever their business, end-users are able to access information at low cost from anywhere in the world, 24 hours a day 365 days a year. And the only investment required is a phone line and PC.

The new technology of the Internet appeals to a market which is fairly new to information and has very low levels of information literacy. Suddenly a variety of new information sources and search tools which had previously been hard to access are available instantly to users. Just as one may wander around a library in wonderment of the knowledge contained within, the proliferation of information on the Internet provides a taster to what may be out there. Curiosity fuels users to seek more and thus an instant addiction is created. The result is the creation of a broad new interactive market place for the dissemination of information. A market with new users, new rules and new ethics.

3. The Internet dilemma

Despite the benefits of the Internet, users face a number of dilemmas:

- *How to find what is actually needed in a sea of information?* The Internet does not have any obvious boundaries to categorise information neatly into relevant areas of interest. Users seeking to download the latest game have as much place on the Internet as the discerning business professional. Each type of user has to jostle for the information that is required from the mass available. There is no central authority that dictates which search tool to use or makes any guarantees that what is required actually exists. Consequently users are frequently overwhelmed by the amount of information.
- *How long will it take to find useful relevant information?* Finding meaningful results can take minutes to hours. The quantity of information is rarely a problem but trawling through the data to measure its relevance is often time consuming. Overloaded telecommunication systems can further slow the delivery rates of information down to a painful trickle.
- *When information is found, how can it be measured for authority and real quality?* Users have no yardstick by which to measure the value of data found on the Internet. Anyone can publish anything on the Web and often make information sound authoritative. If sources are unknown it is difficult to compare the quality or check the true value of the data. Pharmaceutical researchers have noted claims of cures to diseases published on the Web, only to find that the home page has disappeared within six months. Such claims are doubtful unless they are qualified by authoritative medical publications such as *The Lancet* or *New England Journal of Medicine*.
- *How does the user choose which search tool to use from the array available?* Users have a range of different Web crawlers and search engines available for use offering a variety of different searching aids, all of which have to be learnt for optimal use. The preferred search engine is dependent upon personal preference. It can be difficult to know the best search engine to use.
- *How should searches be constructed to be precise?* For users who are unfamiliar with the way data is structured in a database, or indeed how information is indexed, it is very difficult to become truly proficient in searching. For example a user who may be investigating tachycardia in patients treated with the cold cure Contac would expect to be alerted to synonyms such as phenylpropanolamide. The majority of search engines do not have the sophistication to do this automatically.
- *Will retrieved information be presented in a manner that fits the users environment?* Standardisation of formats is important for users to enable the data to be integrated into internal data management systems. Information from the Internet comes in a multitude of different structures and hence is not so easy to integrate.

4. The benefits traditional online services provide

To put the dilemmas of the new Internet user into perspective it is important to explore the benefits of traditional online hosts. Defined as services which provide access to a range of data through proprietary, command driven systems, usually via a modem, traditional online services have provided an information service for many years. The age of the industry has meant searching techniques have been refined and developed through experience. Experience and customer demand have developed the industry so that the provision of data is timely, sophisticated, structured and authoritative.

Traditional online services such as DIALOG and DataStar provide the 'one-stop shop' for information professionals. Information from a wealth of sources — over seven terabytes of data or 20 billion pages of text and images (Ref 4) — is provided through one interface from over 55 million sources. Compared to the Internet, which is estimated to have 20 million unqualified pages (Ref 5), traditional online is infinitely more valuable to the business user. The benefit of cross file searching adds real value to the information by ensuring that all relevant matches are identified as well as removing duplicates. The value of the data is confirmed because all data available is quantified and carefully checked by the database providers, using only high-quality sources. The key benefit given to information supplied through a host is the value added service of sophisticated indexing, enabling users to target the information required precisely in minimal time.

5. Role of information providers

Database producers play an integral part in the development of the Internet. The real value of information lies in its quality. Information providers are key to the authentication of data, ensuring the real value of data by continuing to use established, known and respected sources. Updating of sources is also vital to the value of the data; where appropriate some sources can be updated several times a day. If the value of data is undermined users will never gain faith in the services which they use and pay for. The role of the information providers is important because they are able to be a policing service of sorts to validate the information.

Information providers have traditionally been strong in the structure and content of databases rather than its accessibility. The Internet, however, opens a new channel through which information providers can deliver information. The new technology is beginning to blur the lines of information provision.

Hosts package different information providers' material into one service. The benefit to the user is quite distinct because it enables the user to search for *all* relevant information on a topic, across a number of information providers, at the same time. Standardisation and cross indexing added to information providers data by the hosts is a powerful tool for the user to have at hand. It provides users with sophisticated and precise searching techniques. The added value to information providers is that their information is made available to a wider audience and is made more potent because a selection of quality sources has more impact and value for a user than just a single source.

For the successful future role of information providers it will be important to pursue the following actions:

- continue to concentrate upon the provision of quality authoritative data;
- work with hosts to deliver data to users in a format that they expect and thereby reach a new audience

6. Role of online hosts

Online hosts *add value* to data through standardisation, which enables a number of benefits to be delivered that are not readily available on the Internet, including:

- *one interface to a wide range of data*, enabling users to scan a multitude of sources using the same search term and to deal with one supplier rather than lots of individuals. Many databases can be searched simultaneously and just one of these databases may contain five million individual records. A Web surfing exercise would take hours to achieve an inferior result;
- *multi file searching*, providing the user with the confidence that all relevant information will be selected. Hosts also provide relevancy ranked searches where the host's software analyses occurrences of search terms statistically and provides not just the number of records found but a list of the records which are most relevant to the search terms entered. Current search engines on the Web are not sophisticated enough to pick up all related terms;
- *current awareness*, delivering information automatically, via fax, post and e-mail, the moment that it is published online. Such services provide real value for tracking competitors, alerting users to the latest developments. The user only has to record the search criteria once and then let it run automatically without ever going online; results will be sent automatically. Although alerts are available on the Web they can only be as accurate as the search engines behind them;
- *document delivery* from multiple sources, enabling users to order the original documents directly online. This provides the user with the completely integrated solution. Documents can be sourced from the Web but copyright can be an issue as well as the limited range of documents available;

- *dedicated customer service desks*, available for specialist advice on database content and search guidance. Customer services are available around the clock to answer specific questions about content. Full supporting literature is available for all sources. On the Internet you are on your own.

In a nutshell users can be confident that data retrieved through a host is evaluated, high value information from confirmed sources, and the information can be identified within minutes. There is no comparison to the vast number of free Web sites in terms of content. So why does the Internet continue to be perceived as such a threat to traditional online information services?

7. What HTML means — a single platform

To understand why the Internet is having such a tidal wave effect on the information industry it is necessary to look at what is behind the Internet. What are the bones made of? Why is HTML raved about?

The significance of HTML is its evolution as the first industry standard, the first platform to which all services must adhere if they are to integrate truly with other applications and take advantage of the Internet. Debates on the actual sophistication of HTML as a development platform are prevalent. At this point the actual standard — whether it's HTML, Java or Z39.50 — isn't important. The significance is the convergence of services into a single format. When products are developed to run on a Web server using HTML as the development tool they are instantly accessible by any Internet user. This factor places HTML at the heart of the debate about the demise of the traditional online services because it is the first standard to challenge existing proprietary search languages, even though it is a standard with limited functionality.

Hosts are under threat of technology and the adoption of new working environments. Just as users have become familiar with the Windows working environment, they are now adopting the Internet and intranet environments. The Internet is a natural step in the progression of technology. Players who ignore the tide lose out. It is not necessarily the best who wins but the player who anticipates the market's mood, as was shown in the battle between Betamax and VHS or Apple and Microsoft.

8. How will HTML benefit online hosts?

By understanding the benefits of HTML (or any other equivalent standards), online hosts can take advantage of the Internet and use experience learnt in the development of the traditional online services to create real competitive advantage by offering their services on the Internet. Hosts cannot ignore the facts that:

- the Internet has provided a large new distribution channel which is populated by new untapped markets;
- users on the Internet are frequently people who have identified a need for information but not previously had any means to access information sources easily.

HTML is the key to the market because it provides the mechanism through which to deliver products on the Internet to reach this new audience. For the first time there is a clearer direction for product development. The implications impact the development of products for both the Internet as well as intranets. Hosts can pass on some significant benefits to their customers by making services available via the Web. Services provided over the Internet drastically reduce the investment costs which the user previously had to incur to access information: for example, investment in training to improve searching techniques. Sources of information can be provided in one place through one interface. HTML allows for flexible adaptation of products which users can access instantly rather than having to wait for special upgrades to existing software.

9. Adoption of the Internet standard

When online hosts adopt HTML there are multiple benefits which can be delivered to Internet users. As can be seen, hosts do not have to be excluded from the Internet. The advanced search engines which hosts have developed over the years can be translated to the Internet environment. The users would benefit from tools which perform complex searches automatically without the users having to develop their own search techniques. In addition host services can provide targeted information tools, providing access to specific content groups such as patents. Traditional standards of service would continue to be offered via the Web to different audiences without loss of sophistication.

Knight-Ridder Information has developed products using HTML for the Web, which provide different levels of searching control depending on the requirements of the user:

- *KR ScienceBase*, specifically for the bench scientist, feeds off our database resources yet is still able to combine powerful retrieval with a simplified approach to searching;
- a joint venture between Knight-Ridder Information and CompuServe has made available a complete collection of news sources to CompuServe's consumer online clientele;
- a whole suite of Web-based products is under development in the US and in Europe.

The development process has highlighted some of the considerations which must be made when translating

existing search techniques into the Web environment, but is indicative of a commitment to the Web as the future for the electronic dissemination of information. Those wishing to check progress to date may do so now in the exhibition area.

An important area of education to be tackled is that of pricing. Information produced by labour-intensive means, qualified and indexed, can hardly be given away. A significant charge must be levied and information professionals with an appreciation of the worth of such data have long understood this. However, many Web users must also be convinced that a useful piece of information should have a price-tag. On the other hand, one can understand the repulsion of subscription fees greater than a year's worth of searching. Companies choosing this approach may ultimately regret it.

Hosts need to work in close alliance with the information providers to integrate, synthesise and add value to services, to compete with the multitude of new services. More price flexibility in these partnerships is called for. Gateways from hosts to consumer online services are expected to grow in number.

Predictions are that users on the Internet will rapidly come to appreciate the services that a host can provide and begin to demand a higher concentration of such services, which will undoubtedly lead to an increased level of merger activity.

10. New role of online hosts

As technology develops it is inevitable that the usage of the Internet will continue to rise, as did the usage of motorways, but only if the objectives of users are met. The proliferation of information available on the Internet will continue to grow, giving business users more reason to demand better levels of quality of products available on the Web, to steer a satisfactory route through the chaos. To satisfy the demands of this new group of users, the question arises as to how hosts must change to meet this new challenge.

The companies which provide these host services are having to mutate to stay abreast of external developments now impacting on their prescribed territory. Admittedly, some may be doing so belatedly but, like Microsoft, they have the capability and resources to create new products for a changing marketplace.

Although the type of online searching upon which this conference was originally founded may now be less in favour, the hunger for information increases exponentially and the challenge to continue to provide access to renowned information sources will be met. Hosts are very well aware that the Internet is not simply a threat but a dissemination medium of enormous potential. The past eighteen months have seen a burgeoning of Web and Windows products: DataStar, DIALOG, Profile, MAID and STN have all entered the fray. Inevitably, some will fall wide of the mark and a few will disappear. But, it was ever the case in our industry.

The key to success will be to keep sight of what is the real benefit offered to users, by hosts. That is to search precisely across many data sources, supported by the most comprehensive collection of *authoritative* data provided through one interface. Providing services which add value to the data by synthesising it and delivering it in the format that users require is now crucial: for example, the ability to transfer data automatically into groupware environments such as Lotus Notes. This approach ensures that hosts are able to maintain the advantage by continuing to offer the best collection of data from one source. Partnering with the information providers will ensure that the level of service provided continues to out-perform all other available sources.

11. Conclusion

In short, yes there is evidence for the demise of traditional, command-driven, online services, but they have not reached the end of the road!

The Internet provides the ideal environment which will allow hosts to thrive because for the first time technology is at a stage that will permit hosts to deliver the same searching sophistication that could traditionally only be delivered in command-driven services. The real benefit of the Internet is that travellers who appreciate all that the country lane routes have to offer can continue to appreciate those familiar sites while travelling at speed down the motorway.

The Internet signifies the convergence of two worlds. A rapidly evolving technology is meeting an older more constrained information repository head on. As this evolution continues, customers become more sophisticated in their expectations. The net result will be online information products with a truly different look and feel from those which have been in use for so many years.

Although traditional online hosts may be in decline they are far from becoming extinct. Direct distribution, joint ventures, price deals and intelligent Web navigators are either adverse external pressures or new opportunities in which to be involved and so remodel one's product offerings. It is a certainty that companies with proven longevity will reinvent themselves sufficiently to see in the millennium.

The future for untraditional online hosts is very positive and a future that is worth being excited about.

Karen Thomas
Marketing Manager Europe
Knight-Ridder Information
Haymarket House
1 Oxendon Street
London SW1Y 4EE
UK
Tel: +44 (0)171 930 7646
Fax: +44 (0)171 930 2581
E-mail: karen.thomas@dm.krinfo.ch

References:

- [1] APT Data Services (1996) Share and bond issues: Alta Vista Internet Software Inc., *Computergram International* no.987, 29 August.
- [2] Mowat, M. (1996) Recently established Web sites, *Information World Review*, 115, 26.
- [3] <http://edina.ed.ac.uk/>
- [4] Knight-Ridder Information (1996) *Corporate Capabilities*, Mountain View, California.
- [5] Shein, E. (1996) 'Twas a fine year for webmasters, *PC Week*, 13(26), 39



Author Index

Aguarón, Isaac Bolea	247	Miido, H.	21
Allen, G.	29	Miyakawa, Takayasu	331
Anagnostelis, Betsy	33		
Barberá, José	187	Nicholas, David	343
Barker, Philip	261	Nieuwenhuysen, P.	367
Benz, Joachim	151	Nuyts, Annemie	225
Boelio, David B.	167		
Bothma, Theo J. D.	459	Ochsner, H.	477
Bowman, Benjamin F.	231	O'Riordan, Gobnait	77
Braekevelt, Patrick	433		
Brown, Sheridan	253	Packer, Joan G.	41
Brown, Suzan A.	181	Piggott, Sylvia E.A.	349
Brüggemann, Rainer	151	Pollitt, A. Steven	433
Brunelle, Bette S.	387	Psohlavec, Stanislav	403
Clausen, Helge	115	Read, Tim	49
Cooke, Alison	33	Ritchie, Ian	161
Crane, Dennis J.	397	Romanov, Nickolay	327
Craney, Linda	447	Rose, Tony G.	65
		Ruiz, Maria Jesús Fernández	247
Day, Joan	357	Rusch, Peter F.	221
de Prado, Rosario Lopez	417	Russell, Aileen	471
Delcroix, Jean-Claude	305		
Dixon, Ross	453	Sale, Elizabeth	195
		Scott, T.M.	299
Edwards, Catherine	357	Silhanek, Jaroslav	71
		Smith, Martin P.	433
Fahy, Sarah	125	Snyman, Dorette	85
Fourie, Ina	85	Sodipe, R.O.	293
Frank, Ariel	211	Subba Rao, S.	107
Frossling, Ingrid	343	Swan, Alma	253
Garcia-Sierra, A.J.	203	Thomas, K.	477
Gilbert, Philippe	321	Tomauiuolo, Nicholas G.	41
Grant, C.A.	299	Treglown, Mark	433
Grygierczyk, Natalia	409		
Guittet, Christian	423	Ullate, José Maria Angós	247
Hall, Hazel	49, 471	van Brakel, Pieter A.	57
Hanani, Uri	211	Vanouplines, P.	367
Hey, Jessie	13	Voigt, Kristina	151
Hildyard, C.J.	143		
Hjelm, Johan	283	Waddington, Paul	277
		Wallas, Philip	237
Ihlenfeldt, Wolf-D.	135	Walton, Graham	357
Inbar, David	129	Warner-Allen, Charles	93
		Webb, T.D.	241
Jonckheere, Charles	225	Weiner, Michael L.	221
		Weiss, Arthur	391
Karen, Vladimir	403	Welsh, Sue	379
Kempf, Alois	99	Whitaker, B.J.	143
Kinghorn, Cris	363	White, Martin	271, 313
Klobas, Jane E.	173	Winzenried, Arthur	7
Knight, Nancy H.	167	Worlock, David R.	1
McNab, Alison	33		
Meek, B.	279		

Online Information 96 Proceedings

Page 483



Subject Index

- Academic environments 211,253,271,409,625
Adobe Acrobat 453
Affordability 129
Africa 293
Alta Vista 387
Archives 403
Asia 331
Authoring tools 423,447
- Bandwidth 279,283,305
Book ordering 21
Browsers — see User interfaces
Business information 1,107,125,253,271,363,391
- CD-ROM distribution 327
CD-ROM mastering — see CD-ROM publishing
CD-ROM publishing 161,327,403
CD-ROM use — see Database usage
CD-ROMs 167,239,231,343
Charging — see Pricing strategies
Chemical Abstracts 195
Chemical information 135,143,151,171,195
Chemical structures 135
Chemicals — see Chemical information
Collaborative working 187,299
Collection development .99,211,237,241,409,471
Communication channels 203,253,299
Company information 341,391
Competitive intelligence 181,349,391
Competitor analysis 349,391
Computer graphics 423
Computer networks — see Networks
Computer-mediated communications 173,299
Computerized information services — see
Library/information services,
Home information services
Consumer online services — see Home information
services
Consumers — see End users
Corporate information environments 173,181,
.....187,195,349,471
Cost reduction 129
Costs 305
Course content 57,77,85
Copyright 1
Credit information 391
Current awareness services 195
Customers — see End users
Czech Republic 71,403
- Data communications 261,279,293,299
Data compression 276,403
Data protection 1
Data quality 115,125,151, 187, 379, 477
Database content 271,313
Database descriptions 225,241
Database evaluations — see Evaluations
- Database quality — see Data quality
Database usage 331
Delphi study 7
Denmark 115
Digital libraries — see Electronic libraries
Digitization 187,403,409
Disintermediation 29, 231,357,363,433
Distance education 85
Document creation 143,453
Document delivery 143,225
Dow Jones News/Retrieval 167
DR-LINK 221
- Education and training 29,57,65,71,77,85,
Electronic document delivery — see Document
delivery
Electronic journals 143
Electronic libraries 13,211,241,397,403,409
Electronic products 459
Electronic publishing ...143,161,167,237,397,423
.....459
see also Web publishing
EMBASE 433
End users 21,29,41,231,253,343,357,433,477
Environmental chemicals 151
EPOQUE 225,433
ERNET 107
Europe 305
Evaluations 33,151
see also User studies
Excite 387
Expert systems 211
- Factual databases — see Non-bibliographic
databases
Financial information 93,125,391
Forestry information 99
Fractals 279
France 321
FrontPage 447
Full text 225,271,459
Future studies — see Trends
- GiAS 107
Government information 1,247
Graphical User Interface (GUI) — see User
interfaces
Hasse diagrams 151
Hawaii 241
Health information — see Medical information
HIBROWSE 433
Home information services 313,321
Flynn Home pages — see Web pages
Hot Dog Professional 447
HotJava 49
HoTMetaL Pro 447

- HTML 143,321,423,447,453,477
- Human computer interaction — see also User interfaces
- Hybrid systems 231
- IBM InfoMarket 167
- Image databases — see Image processing
- Image processing 279,403,423,453
- Image transmission 279
- IMPEL2 project 357
- Indexing 379,387
- India 107
- Information access 1,135,173,237,261,253,283,305,403,417
- Information centres — see Library/information centres
- Information delivery 187,253,261,271,279,283,397,459
- Information gathering 135,173,211
- Information industry 161,327,331
- Information management 7,349,363,409,471
- Information markets 161,327,331
- Information needs — see User studies
- Information policy 173,203,363,409
- Information professionals 7,13,29,77,181,241,349,357
- Information providers — see Information provision
- Information provision 7,107,125,173,261,283,321,331,477
- Information quality — see Data quality, Quality control
- Information retrieval 49,57,71,115,135,211,221,231,241,367,379,387,409,417
- Information services — see Library/information services
- Information skills 13,349,433
- Information Society 1,203
- Information sources 71,77,115,125,241,271,293,391,409,471
see also Internet resources
- Information superhighway 65,203,299
see also Internet
- Information systems 211,221,225,231,283
- Information technology 71,93,195,279,283,423
- Information transfer 49
- Information use 253,305
- Intelligent agents 13,203
- Intelligent systems — see Expert systems
- Interactive advertizing — see Advertizing
- Interlibrary loans 21
- Intermediaries — see Information professionals
- Internet resources 1,33,99,107,115,135,151,211,241,247,313,367,379,391,471
- Internet use 77,85,115,143,247,253,299,343
- Internet 49,57,65,107,167,173,187,203,241,293,279,299,305,321,477
see also Information Superhighway, World Wide Web
- Intranets 173,187,195,211,279
- Ireland 77
- Japan 331
- Java 49,367
- KATSIR 211
- Knowbots — see Intelligent agents
- Knowledge databases — see Expert systems
- Knowledge representation — see Expert systems
- Korea 331
- Legal information 1
- Librarians — see Information professionals
- Libraries — see library/information centres
- Library automation 21,225,417
- Library/information centres 181,237,241,417,471
- Library/information services 21,115,181,195,231,241,247,283,331,343,409,417,477
- Life-cycle costs 129
- Lotus Notes 195,247
- Manuscripts 403
- Marketing 27,181,313
- Marketing 93,129,167,271,313
- Medical information 433
- MEDLINE 231
- Memory of the World project 403
- Metadatabases 151,379
- Metered New Issues 167
- Minitel 321
- Mobile information systems 261,283
- Multimedia 85,143,423,459
- Museums 417
- Natural language searching 41,221,397
- Navigation techniques 57,379,459
- Navigator Gold 447
- Netherlands 409
- Networking — see Networks
- Networks 49,107,167,173,187,293,299,305,313,321,417
see also Internet, Intranets
- News information 107,125,161,271,391
- Newspapers — see News information
- NICNET 107
- Nigeria 293
- Notebook computers 261
- Object-oriented systems 49
- OMNI 379
- On the Move project 283
- Online catalogues 21
- Online hosts 477
- Online information retrieval — see Information retrieval
- Online public access catalogues — see Online catalogues
- Online resources — see Information sources, Internet resources
- Online searching 29,41,57,85,135,221,225,231,305,367
- Online services — see Library/information services
- Online usage — see Database usage
- OPACs — see Online catalogues
- Page Mill 447
- Palm/top computers 261
- Patents 225

Payment systems	93,271,313,327	SuperJanet	299
Portable computers	261,283	Supply chain management	129
PRAISE	241		
Preservation	403	Taiwan	331
Pricing strategies	167,271,313	TCP/IP protocol	187
Production process	129,203,423	Teaching — see Education and training	
Psion 3a computers	261	Telecommunications	293,305
Public administrations	247	Teletext	161,313
Publishing	161	Thailand	331
see also CD-ROM publishing, Electronic		Training — see Education and training	
publishing, Web publishing		Trends	7,65,331,357,447
Purchasing	129,203		
		Universities — see Academic environments	
Ranking	151,397,387	User interfaces	21,41,221,225,231,433,459
Regulatory issues	93	User needs — see User studies	
Revenues	313	User studies ...	115,129,173,195,253,305,357,363
Review services	33,247,367,387	~	409,471
Russia	327	Userbots — see Intelligent agents	
		Users — see End users, User studies	
Schools — see Academic environments			
Search engines	41,57,367,387	Value adding	143,459
Search strategies — see Search techniques		Video compression	279,283
Search techniques	41,57,135,387,433	Videotex	161
Searchbots — see Intelligent agents		View-based search systems	433
Selling	181,203	Virtual reality	13
Service providers	107,397		
SGML	143,423,453	Web pages	271,367,447,471
Singapore	331	Web publishing	93,143,161,447,453,459
Softnet	107	Web searching — see Online searching, Search	
Software packages	49	engines	
South Africa	85	World Wide Web (WWW)	41,57,107,135,167,
Spain	417		241,247,299,313,321,367,379,403,447
Statistics	423	see also Internet	
Stock exchanges	93		
Subscription agents	433	Yahoo!	93